

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 63.3034 Seconds
(without alignments)
1067.241 Million cell updates/sec

Title: US-10-087-631B-4
Perfect score: 30
Sequence: 1 tggactcagtccttgatcaccctctt 30

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapept 1.0

Searched: 2185239 segs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20.4	68.0	340	24	ABL70738
C 2	20.4	68.0	1323	21	AAC76076
C 3	20.4	68.0	1449	22	AAH90080
C 4	19.6	65.3	15809	22	AAK68321
C 5	19.6	65.3	81940	22	AA505390
C 6	19.6	65.3	81940	24	ABK64829
C 7	19.4	64.7	300	21	AAA00522
C 8	19.4	64.7	840	20	AAZ27923
C 9	19.4	64.7	840	20	AAZ27924

C 10	19.4	64.7	987	20	AAZ27915
C 11	19.4	64.7	987	20	AAZ27916
C 12	19.4	64.7	1795	20	AAZ27921
C 13	19.4	64.7	1795	20	AAZ27922
C 14	19.4	64.7	1897	20	AAZ27913
C 15	19.4	64.7	1897	20	AAZ27914
C 16	19.4	64.7	1897	20	AAZ27915
C 17	19.4	64.7	1897	20	AAZ27916
C 18	19.4	64.7	1897	20	AAZ27917
C 19	19.4	64.7	1897	20	AAZ27918
C 20	19.4	64.7	1897	20	AAZ27919
C 21	19.4	64.7	1897	20	AAZ27920
C 22	19.4	64.7	1897	20	AAZ27921
C 23	19.4	64.7	1897	20	AAZ27922
C 24	19.4	64.7	1897	20	AAZ27923
C 25	19.4	64.7	1897	20	AAZ27924
C 26	19.4	64.7	1897	20	AAZ27925
C 27	19.4	64.7	1897	20	AAZ27926
C 28	19.4	64.7	1897	20	AAZ27927
C 29	19.4	64.7	1897	20	AAZ27928
C 30	19.4	64.7	1897	20	AAZ27929
C 31	19.4	64.7	1897	20	AAZ27930
C 32	19.4	64.7	1897	20	AAZ27931
C 33	19.4	64.7	1897	20	AAZ27932
C 34	19.4	64.7	1897	20	AAZ27933
C 35	19.4	64.7	1897	20	AAZ27934
C 36	19.4	64.7	1897	20	AAZ27935
C 37	19.4	64.7	1897	20	AAZ27936
C 38	19.4	64.7	1897	20	AAZ27937
C 39	19.4	64.7	1897	20	AAZ27938
C 40	19.4	64.7	1897	20	AAZ27939
C 41	19.4	64.7	1897	20	AAZ27940
C 42	19.4	64.7	1897	20	AAZ27941
C 43	19.4	64.7	1897	20	AAZ27942
C 44	19.4	64.7	1897	20	AAZ27943
C 45	19.4	64.7	1897	20	AAZ27944

ALIGNMENTS

RESULT 1	ABL70738/c	ABL70738 standard; cDNA: 340 BP.
ID	ABL70738	
AC	ABL70738	
XX		
DT	14-MAY-2002 (first entry)	
XX		
DE	Corn tassal-derived polynucleotide (cdps) SEQ ID NO:112.	
XX		
KM	Corn; corn tassal-derived polynucleotide; cdps; hybrid breeding; CDPS;	
KW	Inheritance; characterisitic; growth; development; disease resistance;	
KW	environmental adaptability; quality; yield; molecular marker;	
KW	multigene trait; plant breeding; corn tassal; gene; ss.	
OS	Zea mays.	
XX		
PN	US2001051335-A1.	
XX		
PD	13-DEC-2001.	
XX		
PF	16-APR-1999; 99US-0294093.	
XX		
PR	21-APR-1998; 98US-082567P.	
XX		
PA	(LALG/) TALGUDI R V.	
PA	(ITOL/) ITO L Y.	
PA	(SHER/) SHERMAN B K.	
XX		
PI	Laljudi RV, Ito LY, Sherman BK;	
XX		
DR	WPI: 2002-163647/21.	

Canine B7-2 protei
Complementary str
Canine B7-2s prote
Canine B7-2 gene
Canine B7-2 gene
Canine B7-2 gene c
Nucleotide sequenc
Probe, SK535, for
DNA encoding novel
Human immune/haema
Drosophila melanog
Human musculoskele
Human immune/haema
Plant microsatelli
Rat sequence diffe
DNA encoding novel
Human gamma-synuci
Human brain expres
Human gamma synuci
Human secreted pro
Human breast cance
Human HBBBA67A DNA
Breast cancer rela
Human persyn gene.
Human polynucleoti
Human breast speci
Human breast and o
Streptococcus pneu
C glutamicum codin
DNA encoding a S.
Streptococcus pneu
DNA encoding S. pn
Streptococcus pneu
Streptococcus pneu
Breast cancer rela
Breast cancer rela

XX Novel purified corn tassel-derived polynucleotide useful for
 PT determining altered gene expression, to recover regulatory elements and
 PT to follow inheritance of desirable characteristics through hybrid
 PT breeding programs

PS Claim 1; SEQ ID 112; 201bp; English.

XX The present sequence describes a purified corn tassel-derived
 CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
 CC selected from those given in ABL70627 to ABL76833. The cdps sequences
 CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I)
 CC can be used for determining altered gene expression, to recover
 CC regulatory elements and to follow inheritance of desirable
 CC characteristics through hybrid breeding programs. (I) are also useful
 CC in the evaluation, and alteration of desired characteristics associated
 CC with growth and development, disease resistance, environmental
 CC adaptability, quality and yield, and as molecular markers for studying
 CC inheritance of multigenic traits in a plant breeding program. (I) can be
 CC used to produce a tassel-specific profile of gene transcription, a
 CC transcript image, to clone regulatory elements for use in transformation
 CC vectors, to express a polypeptide, to identify, isolate or extend
 CC identical or related corn tassel nucleic acid sequences from DNA
 CC libraries, in nucleic acid hybridization or amplification technologies,
 CC as query sequences to determine homology of known sequences, as probe
 CC for use in Southern or Northern hybridisation, and to identify the
 CC presence of and/or to determine the degree of similarity between two
 CC (or more) nucleic acid sequences

SQ Sequence 340 BP; 102 A; 52 C; 84 G; 89 T; 13 other;

Query Match

Best Local Similarity 68.0%; Score 20.4; DB 24; Length 340;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 1 TGGACTCAGCTCGTCATCTCACTCTCT 30
 126 TGGACACGCTACATGTTATCTCATCTTCT -97

RESULT 2

AACT6076 standard; cDNA; 1323 BP.

AACT6076;

08-FEB-2001 (first entry)

Human ORF1631 polynucleotide sequence SEQ ID NO:3261.

Human; open reading frame; ORF; detection; cytostatic; hepatotropic;
 human; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
 anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiant;
 immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 hypotensive; dermatological; immunosuppressive; antineoplastic;
 antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
 antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 neurodegenerative disorder; osteoarthritis; graft vs host disease;
 cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 cholesterol ester storage; systemic lupus erythematosus; infection;
 severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 allergic; aplastic anemia; nocturnal haemoglobinuria; burn; wound;
 thrombosis; contraceptive; ss.

Homo sapiens.

MO200056473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shinkens RA, Leach M;

WPI, 2000-602362/57.

P-PSDB; AAB41867.

Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease

Claim 5; Page 2462; 5507bp; English.

AACT74446 to AACT7606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORF open reading frames 1 to 3161. The ORF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiparkinsonian; neurotrophic; neuroprotective;
 CC osteopathic; anticonvulsant; antidiabetic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antirheumatic; antiviral; antihypertensive;
 CC antihypertensive; antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORF-associated disorder. The
 CC nucleic acids can be used to express ORF proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergic, aplastic anemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antineoplastic disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

SQ Sequence 1323 BP; 239 A; 435 C; 343 G; 304 T; 2 other;

Query Match

Best Local Similarity 68.0%; Score 20.4; DB 21; Length 1323;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 1 TGGACTCAGCTCGTCATCTCACTCTCT 30
 997 TGGCTCAGCTCGTCATCTCACTCTCT 1026

RESULT 3

AAH90080/c standard; cDNA; 1449 BP.

AAH90080;

01-OCT-2001 (first entry)

Human bone marrow cDNA, SEQ ID NO: 324.

Human; bone marrow; antineoplastic; cytostatic; neuroprotective;
 human; antiparkinsonian; neurotrophic; hepatotropic; vulnary;
 immunosuppressive; gene therapy; cytokine cell proliferation;
 cell differentiation modulator; immune disorder; infection; cancer;
 human immunodeficiency virus; HIV; autoimmune disorder; hemophilia, ss.

Homo sapiens.

WO200153453-A2.

26-JUL-2001.

PF 23-DEC-2000; 2000WO-US34960.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0596042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 30-NOV-2000; 2000US-0250583.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;
PI Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao Qa, Zhou P, Drmanac RT;
XX
DR WPI: 2001-488707/53.
DR P-PSDB; AAM00961.
XX
PT Novel bone-marrow-expressed polynucleotides and polypeptides, useful
PT for treating e.g. cancer and immune deficiency disorders -
XX
PS Claim 1; Page 429; 648pp; English.
XX
CC The present sequence is one of 251 novel human polynucleotides
CC expressed in the bone marrow. The polynucleotide and the
CC polypeptide encoded by it are useful in the treatment of various
CC immune deficiencies and disorders. The deficiencies and disorders may
CC be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal
CC infection, or may result from an autoimmune disorder, a coagulation
CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,
CC suppression of an inflammatory response or treatment of a nervous
CC system disorder such as Alzheimer's disease. Detection of the presence
CC or increased expression of the polynucleotide or the protein it
CC encodes is useful for the diagnosis and/or prognosis of one
CC or more types of cancer. The polynucleotide and polypeptide can be
CC used as nutritional sources or supplements and in the screening of
CC chemical compounds as potential drugs.
XX
SQ Sequence 1449 BP; 334 A; 394 C; 446 G; 274 T; 1 other;
Query Match 68.0%; Score 20.4; DB 22; Length 1449;
Best Local Similarity 80.0%; Pred. No. 22;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 TGGACTCAGTCTTGGTCAATCTACCTTCT 30
DB 327 TGGTCTCACTCTTGGACTTGTACCTTGT 298
RESULT 4
AAK68321
ID AAK68321 standard; DNA; 15809 BP.
XX
AC AAK68321;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23133.
XX
KW Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189674.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 20-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.
 PR 08-NOV-2000; 2000US-0244675.
 PR 08-NOV-2000; 2000US-0244676.
 PR 08-NOV-2000; 2000US-0244677.
 PR 08-NOV-2000; 2000US-0244678.
 PR 08-NOV-2000; 2000US-024523.
 PR 08-NOV-2000; 2000US-024524.
 PR 08-NOV-2000; 2000US-024525.
 PR 08-NOV-2000; 2000US-024526.
 PR 08-NOV-2000; 2000US-024527.
 PR 08-NOV-2000; 2000US-024528.
 PR 08-NOV-2000; 2000US-024552.
 PR 08-NOV-2000; 2000US-0246509.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0250160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMA GENOME SCI. INC.
 XX
 PI Rosen CA; Barash SC, Ruben SM;
 XX
 DR WPI; 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX

PS Disclosure; SEQ ID NO 23133; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytosolic
 CC activity and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 15809 BP; 3658 A; 4263 C; 4050 G; 3838 T; 0 other;

Query Match 65.3%; Score 19.6; DB 22; Length 15809;
 Best Local Similarity 84.6%; Pred. No. 78;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 DY 3 GACTCAGTCTTGCTCATCTGCACCTT 28
 DB 13701 GACTCAGTCTTGCTCATCTGCCTT 13726

RESULT 5
 ID AAS05390 standard; DNA; 81940 BP.
 XX

AC AAS05390;
 DT 24-OCT-2001 (first entry)
 XX
 DE Human titin (connectin) gene sequence.

KW Human; titin; connectin; pickwick mutation; cardiac specific exon; NZB;
 KW titin-related disease; zebrafish; heart failure; heart disease; ds.
 XX
 OS Homo sapiens.

FH Key Location/Qualifiers
 FT CDS 133..80913
 FT /*tag= a
 FT /product= "titin"
 PN WO200151666-A1;

PD 19-JUL-2001.
 PF 12-JAN-2001; 2001WO-US01212.
 XX
 PR 12-JAN-2000; 2000US-01175787.

PA (GEHO) GEN HOSPITAL CORP.
 PI Fishman MC;
 XX
 DR WPI; 2001-451869/48.
 DR P-PSDB; AAD05396.

PT Determining if a subject has or is at risk of developing a
 PT titin-related disease or condition, particularly heart failures,
 XX comprises detecting the presence of a mutation in the titin gene
 PS Disclosure; Page 35-57; 114pp; English.

CC The present sequence encoding for human titin (also known as connectin)

[illegible]


```

XX AC AA227915;
XX XX 20-DEC-1999 (first entry)
XX XX
DE Canine B7-2 protein coding sequence.
XX XX
XX B7: CTLA4; T cell costimulatory protein; dog; cat; autoimmune disease;
XX KM allergic reaction; infectious disease; tumor development; canine;
XX KM graft rejection; inflammation; arthritis; atopic dermatitis; ss.
XX OS
XX OS Canis familiaris.
XX PN WO9947558-A2.
XX PD 23-SEP-1999.
XX PF 19-MAR-1999; 99WO-US06187.
XX PR 19-MAR-1998; 98US-0078765.
XX PR 17-APR-1998; 98US-0062597.
XX PA (HESK-) HESKA CORP.
XX PI Slim G, Yang S, Sellins KS;
XX PI WPI: 1999-571822/48.
XX DR P-PSDB; AAY41076.
XX PT New isolated B7 and CTLA4 nucleic acids, used to develop products for
XX PT treating, e.g. autoimmune and atopic diseases
XX PS Claim 1; Page 102-103; 148pp; English.
XX XX
XX CC The invention provides B7 and CTLA4 (T cell costimulatory proteins)
XX CC encoding nucleic acid molecules from dogs and cats. The proteins can be
XX CC expressed by standard recombinant methodology. The nucleic acid molecules
XX CC and the encoded proteins can be used for preventing or treating diseases,
XX CC e.g. autoimmune diseases, allergic reactions, infectious diseases, tumor
XX CC development, graft rejection, inflammation, arthritis and atopic diseases
XX CC such as atopic dermatitis. They can be used in mammals such humans, dogs,
XX CC cats, cattle, sheep or pets. The products can also be used for detection,
XX CC diagnosis and drug screening.
XX SQ Sequence 987 BP; 315 A; 215 C; 204 G; 253 T; 0 other;

Query Match 64.7%; Score 19.4; DB 20; Length 987;
Best Local Similarity 79.3%; Pred. No. 57;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GGACTCAGTCTTGTCATCTCACCCTTCT 30
DB | |||| | |||| | |||| | |||| | ||||
901 GACTCTTCTCTGCTGCTCCTCCTCTCTCT 873

RESULT 11
AA227916
ID AA227916 standard; DNA; 987 BP.
XX
XX AA227916;
XX XX
XX 20-DEC-1999 (first entry)
XX DE Complementary strand of canine B7-2 coding sequence.
XX XX
XX B7: CTLA4; T cell costimulatory protein; dog; cat; autoimmune disease;
XX KM allergic reaction; infectious disease; tumor development; canine;
XX KM graft rejection; inflammation; arthritis; atopic dermatitis; ss.
XX OS
XX OS Canis familiaris.
XX PN WO9947558-A2.
XX XX

```

```

PD 23-SEP-1999.
XX PF 19-MAR-1999; 99WO-US06187.
XX PR 19-MAR-1998; 98US-0078765.
XX PR 17-APR-1998; 98US-0062597.
XX PA (HESK-) HESKA CORP.
XX PI Slim G, Yang S, Sellins KS;
XX PI WPI: 1999-571822/48.
XX DR P-PSDB; AAY41076.
XX PT New isolated B7 and CTLA4 nucleic acids, used to develop products for
XX PT treating, e.g. autoimmune and atopic diseases
XX PS Claim 1; Page 103-104; 148pp; English.
XX XX
XX CC The invention provides B7 and CTLA4 (T cell costimulatory proteins)
XX CC encoding nucleic acid molecules from dogs and cats. The proteins can be
XX CC expressed by standard recombinant methodology. The nucleic acid molecules
XX CC and the encoded proteins can be used for preventing or treating diseases,
XX CC e.g. autoimmune diseases, allergic reactions, infectious diseases, tumor
XX CC development, graft rejection, inflammation, arthritis and atopic diseases
XX CC such as atopic dermatitis. They can be used in mammals such humans, dogs,
XX CC cats, cattle, sheep or pets. The products can also be used for detection,
XX CC diagnosis and drug screening.
XX SQ Sequence 987 BP; 253 A; 204 C; 215 G; 315 T; 0 other;

Query Match 64.7%; Score 19.4; DB 20; Length 987;
Best Local Similarity 79.3%; Pred. No. 57;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GGACTCAGTCTTGTCATCTCACCCTTCT 30
DB | |||| | |||| | |||| | |||| | ||||
87 GACTCTTCTCTGCTGCTCCTCCTCTCTCT 115

RESULT 12
AA227921/C
ID AA227921 standard; DNA; 1795 BP.
XX
XX AA227921;
XX XX
XX 20-DEC-1999 (first entry)
XX DE Canine B7-2S protein encoding DNA.
XX XX
XX B7: CTLA4; T cell costimulatory protein; dog; cat; autoimmune disease;
XX KM allergic reaction; infectious disease; tumor development; canine;
XX KM graft rejection; inflammation; arthritis; atopic dermatitis; ss.
XX OS
XX OS Canis familiaris.
XX PN WO9947558-A2.
XX PD 23-SEP-1999.
XX PF 19-MAR-1999; 99WO-US06187.
XX PR 19-MAR-1998; 98US-0078765.
XX PR 17-APR-1998; 98US-0062597.
XX PA (HESK-) HESKA CORP.
XX PI Slim G, Yang S, Sellins KS;
XX PI WPI: 1999-571822/48.
XX DR P-PSDB; AAY41076.
XX PT New isolated B7 and CTLA4 nucleic acids, used to develop products for
XX PT treating, e.g. autoimmune and atopic diseases

```

XX Claim 1; Page 109-111; 148bp; English.
 PS
 CC The invention provides B7 and CTLA4 (T cell costimulatory proteins)
 CC encoding nucleic acid molecules from dogs and cats. The proteins can be
 CC expressed by standard recombinant methodology. The nucleic acid molecules
 CC and the encoded proteins can be used for preventing or treating diseases,
 CC e.g. autoimmune diseases, allergic reactions, infectious diseases, tumor
 CC development, graft rejection, inflammation, arthritic and atopic diseases
 CC such as atopic dermatitis. They can be used in mammals such humans, dogs,
 CC cats, cattle, sheep or pets. The products can also be used for detection,
 CC diagnosis and drug screening.
 XX
 SQ Sequence 1795 BP; 592 A; 366 C; 347 G; 490 T; 0 other;

Query Match
 Best Local Similarity 64.7%; Score 19.4; DB 20; Length 1795;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GGACTGAGTCTTGGTATCAGCTTCT 30
 DB 760 GTACCTTCTTGGTATCAGCTTCT 732

RESULT 13

AA227922 ID AA227922 standard; DNA; 1795 BP.

AA227922; AC

20-DEC-1999 (first entry); DT

Canine B7-2S gene complementary DNA sequence. DE

B7; CTLA4; T cell costimulatory protein; dog; cat; autoimmune disease;
 allergic reaction; infectious disease; tumor development; canine;
 graft rejection; inflammation; arthritis; atopic dermatitis; ss.
 OS
 XX Canis familiaris.

PN WO947558-A2.

23-SEP-1999. PD

19-MAR-1999; 99WO-US06187. PF

19-MAR-1998; 98US-0078765. PR

17-APR-1998; 98US-0062597. PR

(HESK-) HESKA CORP. PA

Slm G, Yang S, Sellins KS; PI

WPI, 1999-571822/48. DR

New isolated B7 and CTLA4 nucleic acids, used to develop products for
 treating, e.g. autoimmune and atopic diseases PT

Claim 1; Page 112-114; 148bp; English. PS

The invention provides B7 and CTLA4 (T cell costimulatory proteins)
 encoding nucleic acid molecules from dogs and cats. The proteins can be
 expressed by standard recombinant methodology. The nucleic acid molecules
 and the encoded proteins can be used for preventing or treating diseases,
 e.g. autoimmune diseases, allergic reactions, infectious diseases, tumor
 development, graft rejection, inflammation, arthritic and atopic diseases
 such as atopic dermatitis. They can be used in mammals such humans, dogs,
 cats, cattle, sheep or pets. The products can also be used for detection,
 diagnosis and drug screening.

Sequence 1795 BP; 490 A; 347 C; 366 G; 592 T; 0 other;

Query Match 64.7%; Score 19.4; DB 20; Length 1795;

Best Local Similarity 79.3%; Pred. No. 64;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GGACTGAGTCTTGGTATCAGCTTCT 30
 DB 1036 GTACCTTCTTGGTATCAGCTTCT 1064

RESULT 14

AA227913/C ID AA227913 standard; DNA; 1897 BP.

AA227913; AC

20-DEC-1999 (first entry); DT

Canine B7-2 protein encoding DNA. DE

B7; CTLA4; T cell costimulatory protein; dog; cat; autoimmune disease;
 allergic reaction; infectious disease; tumor development; canine;
 graft rejection; inflammation; arthritis; atopic dermatitis; ss.
 OS
 XX Canis familiaris.

PN WO947558-A2.

23-SEP-1999. PD

19-MAR-1999; 99WO-US06187. PF

19-MAR-1998; 98US-0078765. PR

17-APR-1998; 98US-0062597. PR

(HESK-) HESKA CORP. PA

Slm G, Yang S, Sellins KS; PI

WPI, 1999-571822/48. DR

New isolated B7 and CTLA4 nucleic acids, used to develop products for
 treating, e.g. autoimmune and atopic diseases PT

Claim 1; Page 97-99; 148bp; English. PS

The invention provides B7 and CTLA4 (T cell costimulatory proteins)
 encoding nucleic acid molecules from dogs and cats. The proteins can be
 expressed by standard recombinant methodology. The nucleic acid molecules
 and the encoded proteins can be used for preventing or treating diseases,
 e.g. autoimmune diseases, allergic reactions, infectious diseases, tumor
 development, graft rejection, inflammation, arthritic and atopic diseases
 such as atopic dermatitis. They can be used in mammals such humans, dogs,
 cats, cattle, sheep or pets. The products can also be used for detection,
 diagnosis and drug screening.

Sequence 1897 BP; 585 A; 400 C; 383 G; 529 T; 0 other;

Query Match 64.7%; Score 19.4; DB 20; Length 1897;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GGACTGAGTCTTGGTATCAGCTTCT 30
 DB 906 GTACCTTCTTGGTATCAGCTTCT 878

RESULT 15

AA227914 ID AA227914 standard; DNA; 1897 BP.

AA227914; AC

20-DEC-1999 (first entry); DT

XX	Canine B7-2 gene complementary DNA sequence.
DE	
XX	B7; CTLA4; T cell costimulatory protein; dog; cat; autoimmune disease;
KW	allergic reaction; infectious disease; tumor development; canine;
KW	graft rejection; inflammation; arthritis; atopic dermatitis; ss.
XX	
OS	Canis familiaris.
PN	W039947558-A2.
XX	
PD	23-SEP-1999.
XX	
PE	19-MAR-1999; 99WO-US06187.
XX	
PR	19-MAR-1998; 98US-0078765.
PR	17-APR-1998; 98US-0062597.
XX	
PA	(HESK-) HESKA CORP.
XX	
PI	Slm G, Yang S, Sellins KS;
XX	
DR	WPI; 1999-571822/48.
PT	
PT	New isolated B7 and CTLA4 nucleic acids, used to develop products for
XX	treating, e.g. autoimmune and atopic diseases.
XX	
PS	Claim 1; Page 101-102; 148pp; English.
XX	
CC	The invention provides B7 and CTLA4 (T cell costimulatory proteins)
CC	encoding nucleic acid molecules from dogs and cats. The proteins can be
CC	expressed by standard recombinant methodology. The nucleic acid molecules
CC	and the encoded proteins can be used for preventing or treating diseases,
CC	e.g. autoimmune diseases, allergic reactions, infectious diseases, tumor
CC	development, graft rejection, inflammation, arthritic and atopic diseases
CC	such as atopic dermatitis. They can be used in mammals such humans, dogs,
CC	cats, cattle, sheep or pets. The products can also be used for detection,
CC	diagnosis and drug screening.
XX	
XX	
SO	Sequence 1897 BP; 529 A; 383 C; 400 G; 585 T; 0 other.
QY	Query Match 64.7%; Score 19.4; DB 20; Length 1897;
	Best Local Similarity 79.3%; Pred. No. 65;
	Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
DB	2 GGACTCAGTCTTGTCATCTCACCCTCT 30
	992 GTACTCTTTCCTTGCTGCTGCACCTTCT 1020

Search completed: July 10, 2003, 19:52:36
Job time : 66.3034 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 ; Search time 13.4157 Seconds
(without alignments)
685.785 Million cell updates/sec

Title: US-10-087-631b-4

Perfect score: 30
Sequence: 1 tggactcagctcttggtatctacacctct 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA:*

1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PTCUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19	63.3	33	1	US-08-299-682-13
C 2	19	63.3	33	1	US-08-347-657-1
C 3	18.4	61.3	550	3	US-08-705-771-1
C 4	18.4	61.3	786	5	PCR-US95-08295-1
C 5	18.4	61.3	1260	4	US-09-081-689-3
C 6	18.4	61.3	1297	4	US-08-858-207A-66
C 7	18.4	61.3	1329	4	US-09-305-984-13
C 8	18.4	61.3	1329	4	US-09-073-541A-13
C 9	18.4	61.3	1526	4	US-09-081-689-1
C 10	18.4	61.3	7881	2	US-08-751-189-1
C 11	18.4	61.3	7881	2	US-09-060-836-1
C 12	18.4	61.3	7881	2	US-09-184-445-1
C 13	18.4	61.3	8900	4	US-09-305-984-25
C 14	18.4	61.3	8900	4	US-09-073-541A-25
C 15	18.4	61.3	28882	4	US-08-961-527-140
C 16	17.8	59.3	3757	2	US-09-016-366A-13
C 17	17.8	59.3	3757	2	US-08-978-404B-19
C 18	17.6	58.7	2101	2	US-08-860-150-1
C 19	17.6	58.7	2101	2	US-09-338-132-1
C 20	17.6	58.7	5406	1	US-07-813-593-3
C 21	17.6	58.7	5406	1	US-07-977-451-5
C 22	17.6	58.7	5406	1	US-07-946-507-3
C 23	17.6	58.7	5406	1	US-08-252-517-5
C 24	17.6	58.7	5406	1	US-07-906-397A-5
C 25	17.6	58.7	5406	1	US-08-601-891-5
C 26	17.6	58.7	5406	2	US-09-021-324-5
C 27	17.6	58.7	5406	2	PCR-US92-02750-7

C 28	17.6	58.7	5406	5	PCR-US92-05401-5	Sequence 5, Appl1
C 29	17.6	58.7	5406	5	PCR-US92-09893-5	Sequence 5, Appl1
C 30	17.6	58.7	5470	2	US-08-443-861-1	Sequence 1, Appl1
C 31	17.6	58.7	5470	4	US-08-193-829B-1	Sequence 1, Appl1
C 32	17.6	58.7	43804	4	US-09-171-461-1	Sequence 1, Appl1
C 33	17.4	58.0	2533	4	US-09-149-922-5	Sequence 5, Appl1
C 34	17.4	58.0	2589	6	5212286-1	Sequence 5, Appl1
C 35	17.4	58.0	4149	2	US-08-737-715-1	Sequence 1, Appl1
C 36	17.4	58.0	35081	2	US-08-752-760A-1	Sequence 1, Appl1
C 37	17.2	57.3	330	1	US-08-081-539-84	Sequence 84, Appl1
C 38	17.2	57.3	330	1	US-08-466-647-84	Sequence 84, Appl1
C 39	17.2	57.3	2347	5	PCR-US96-03965-1	Sequence 1, Appl1
C 40	17.2	57.3	3776	1	US-08-162-809-7	Sequence 7, Appl1
C 41	17.2	57.3	6585	3	US-08-746-111-4	Sequence 4, Appl1
C 42	17.2	56.7	999	1	US-08-416-478A-4	Sequence 4, Appl1
C 43	17.2	56.7	999	2	US-08-474-988B-4	Sequence 4, Appl1
C 44	17.2	56.7	999	2	US-08-394-442B-4	Sequence 4, Appl1
C 45	16.8	56.0	426	4	US-09-712-016-32	Sequence 32, Appl1

ALIGNMENTS

RESULT 1
US-08-299-682-13/C
; Sequence 13, Application US/08299682
; Patent No. 5491063
GENERAL INFORMATION:
APPLICANT: Fisher, Mary Ellen
APPLICANT: Watson, Robert Malcom
TITLE OF INVENTION: Methods for In-Solution Quenching of
TITLE OF INVENTION: Fluorescently Labeled Oligonucleotide Probes
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,682
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Petry, Douglas A.
REGISTRATION NUMBER: 35,321
REFERENCE/DOCKET NUMBER: 9002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2974
TELEFAX: (510) 814-2977
INFORMATION FOR SEQ. ID NO.: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-299-682-13

Query Match 63.3%; Score 19; DB 1; Length 33;
Best Local Similarity 96.8%; Pred. No. 9.2;
Matches 30; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 TGGACTCAGCTCTTGGTATCTACCTCTCT 30
DB 31 TGGACTCAGCTCTTGGTATCTACCTCTCT 1

RESULT 2

US-08-347-657-1/c

Sequence 1, Application US/08347657

Patent No. 5571673

GENERAL INFORMATION:

APPLICANT: Plicone, Teresa

TITLE OF INVENTION: Methods for In-Solution Quenching of

NUMBER OF INVENTION: Fluorescently Labeled Oligonucleotide Probes

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: U.S.A.

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION NUMBER: US/08/347,657

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Pety, Douglas A.

REGISTRATION NUMBER: 35,321

REFERENCE/DOCKET NUMBER: 9037

TELEPHONE: (510) 814-2974

TELEFAX: (510) 814-2977

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 33 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-347-657-1

Query Match

Best Local Similarity 63.3% Score 19; DB 1; Length 33;

Matches 30; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

Db 31 TGACTCAGCTCTTGCTCATCTCAGCTCTCT 1

SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/705,771

FILING DATE: August 30, 1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: MULLINS, J.G.

REGISTRATION NUMBER: 33,073

REFERENCE/DOCKET NUMBER: 325800-346 (PFI96)

TELEPHONE: 973-994-1700

TELEFAX: 973-994-1744

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 550 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-705-771-1

Query Match

Best Local Similarity 61.3% Score 18.4; DB 3; Length 550;

Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

Db 122 GACCCCTCCTGCTCTTCACCTGCT 95

RESULT 5

US-09-081-689-3/c

Sequence 3, Application US/09081689

Patent No. 6155992

GENERAL INFORMATION:

APPLICANT: Wallis, Nicola G.

TITLE OF INVENTION: Breast Specific Genes and Proteins

NUMBER OF INVENTION: 30

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: Throup, John

STREET: 205 GACCCCTCCTGCTCTTCACCTGCT

CITY: Sanjoy

STATE: NJ

COUNTRY: U.S.A.

ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: IBM PC compatible

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION NUMBER: PCT/US95/08295

FILING DATE: 30-JUN-1995

CLASSIFICATION:

SEQUENCE CHARACTERISTICS:

LENGTH: 786 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-081-689-3/c

Query Match

Best Local Similarity 61.3% Score 18.4; DB 5; Length 786;

Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

Db 205 GACCCCTCCTGCTCTTCACCTGCT 178

TITLE OF INVENTION: Histidine Kinase
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2793
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/081,689
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/048,347
FILING DATE: 30-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: G410009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1260 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-081-689-3

Query Match 61.3%; Score 18.4; DB 4; Length 1260;
Best Local Similarity 95.0%; Pred. No. 27;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 GTCCTGGTCATCTCACCCT 28
||||| ||||||| |||||||
Db 792 GTCCTGGTCATCTCACCCT 773

RESULT 6
US-08-858-207A-66/c
Sequence 66, Application US/08858207A
Patent No. 6348328
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: No. 6348328el Compounds
NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A
FILING DATE: 09-MAY-1997

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 1297 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-858-207A-66

Query Match 61.3%; Score 18.4; DB 4; Length 1297;
Best Local Similarity 95.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 GTCCTGGTCATCTCACCCT 28
||||| ||||||| |||||||
Db 792 GTCCTGGTCATCTCACCCT 773

RESULT 7
US-09-305-984-13/c
Sequence 13, Application US/09305984B
Patent No. 6331407
GENERAL INFORMATION:
APPLICANT: Toumanen, Elaine
TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
FILE REFERENCE: 1340-1-016N1
CURRENT APPLICATION NUMBER: US/09/305,984B
CURRENT FILING DATE: 1999-05-05
EARLIER APPLICATION NUMBER: 60/084,399
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: 09/305,984
EARLIER FILING DATE: 1999-05-05
NUMBER OF SEQ ID NOS: 76
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 1329
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-305-984-13

Query Match 61.3%; Score 18.4; DB 4; Length 1329;
Best Local Similarity 95.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 GTCCTGGTCATCTCACCCT 28
||||| ||||||| |||||||
Db 264 GTCCTGGTCATCTCACCCT 245

RESULT 8
US-09-073-541A-13/c
Sequence 13, Application US/09073541A
Patent No. 6448224
GENERAL INFORMATION:
APPLICANT: Toumanen, Elaine
TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
FILE REFERENCE: 1340-1-016
CURRENT APPLICATION NUMBER: US/09/073,541A
CURRENT FILING DATE: 1998-05-06
NUMBER OF SEQ ID NOS: 46

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 1329
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-073-541A-13

Query Match 61.3%; Score 18.4; DB 4; Length 1329;
Best Local Similarity 95.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 GTCTGTGTCATCTCACCCT 28
||||| |||||||
DB 264 GTCTGTGTCATCTCACCCT 245

RESULT 9
US-09-081-689-1/c
Sequence 1, Application US/09081689
Patent No. 6165992

GENERAL INFORMATION:

APPLICANT: Wallis, Nicola G.

APPLICANT: Zalacain, Magdalena

APPLICANT: Throup, John

APPLICANT: Biswas, Sanjoy

TITLE OF INVENTION: Histidine Kinase

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert, Price & Rhoads

STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103-2793

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: FASTSEQ for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/081,689

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/046,347

FILING DATE: 30-MAY-1997

ATTORNEY/AGENT INFORMATION:

NAME: Falk, Stephen T

REGISTRATION NUMBER: 36,795

REFERENCE/DOCKET NUMBER: GM10009

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-994-2488

TELEFAX: 215-994-2222

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1526 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-081-689-1

Query Match 61.3%; Score 18.4; DB 4; Length 1526;
Best Local Similarity 95.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 GTCTGTGTCATCTCACCCT 28
||||| |||||||
DB 461 GTCTGTGTCATCTCACCCT 442

RESULT 10
US-08-751-189-1

Sequence 1, Application US/08751189
Patent No. 5919656

GENERAL INFORMATION:

APPLICANT: Harrington, Lea A.

APPLICANT: Robinson, Murray O.

TITLE OF INVENTION: No. 5919656el Genes Encoding Telomerase Protein

TITLE OF INVENTION: 1

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Amgen, Inc.

STREET: 1840 De Havilland Drive

CITY: Thousand Oaks

STATE: California

COUNTRY: USA

ZIP: 91320-1789

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/751,189

FILING DATE: 15-NOV-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Oleski, Nancy A.

REGISTRATION NUMBER: 34,688

REFERENCE/DOCKET NUMBER: A-433

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 7881 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-751-189-1

Query Match 61.3%; Score 18.4; DB 2; Length 7881;
Best Local Similarity 78.6%; Pred. No. 35;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GGACGACGCTGTCATCTCACCCTC 29
||||| ||||||| ||| ||| |||||
DB 5358 GGACGACGCTGTCATCTCACCCTC 5385

RESULT 11

US-09-060-836-1

Sequence 1, Application US/09060836

Patent No. 5981707

GENERAL INFORMATION:

APPLICANT: Harrington, Lea A.

APPLICANT: Robinson, Murray O.

TITLE OF INVENTION: No. 5981707el Genes Encoding Telomerase Protein

TITLE OF INVENTION: 1

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Amgen, Inc.

STREET: 1840 De Havilland Drive

CITY: Thousand Oaks

STATE: California

COUNTRY: USA

ZIP: 91320-1789

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/060,836

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/751,189
FILING DATE: 15-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Oleski, Nancy A.
REGISTRATION NUMBER: 34,688
REFERENCE/DOCKET NUMBER: A-433
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7881 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-060-836-1

Query Match 61.3%; Score 18.4; DB 2; Length 7881;
Best Local Similarity 78.6%; Pred. No. 35;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GGACTCAGTCTGTCATCTCACCCTTC 29
|||||
DB 5358 GGACACAGTCCGTGGCAGCTGGCCTTC 5385

RESULT 12
US-09-184-445-1
Sequence 1, Application US/09184445
Patent No. 6174703

GENERAL INFORMATION:
APPLICANT: Harrington, Lea A.
APPLICANT: Robinson, Murray O.
TITLE OF INVENTION: No. 6174703el Genes Encoding Telomerase Protein
TITLE OF INVENTION: 1
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen, Inc.
STREET: 1840 De Havilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/184,445
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/751,189
FILING DATE: 15-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Oleski, Nancy A.
REGISTRATION NUMBER: 34,688
REFERENCE/DOCKET NUMBER: A-433
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7881 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-184-445-1

Query Match 61.3%; Score 18.4; DB 4; Length 7881;
Best Local Similarity 78.6%; Pred. No. 35;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GGACTCAGTCTGTCATCTCACCCTTC 29
|||||
DB 5358 GGACACAGTCCGTGGCAGCTGGCCTTC 5385

RESULT 13
US-09-305-984-25/c
Sequence 25, Application US/09305984B
Patent No. 6331407
GENERAL INFORMATION:
APPLICANT: Toumanen, Elaine
TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
FILE REFERENCE: 1340-1-016N1
CURRENT FILING DATE: 1999-05-05
EARLIER APPLICATION NUMBER: 60/084,399
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: 09/305,984
EARLIER FILING DATE: 1999-05-05
NUMBER OF SEQ ID NOS: 76
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 25
LENGTH: 8900
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-305-984-25

Query Match 61.3%; Score 18.4; DB 4; Length 8900;
Best Local Similarity 95.0%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 GTCCCTGGTCATCTCACCCTT 28
|||||
DB 7586 GTCCCGGTGTCATCTCACCCTT 7567

RESULT 14
US-09-073-541A-25/c
Sequence 25, Application US/09073541A
Patent No. 6448224
GENERAL INFORMATION:
APPLICANT: Toumanen, Elaine
TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
FILE REFERENCE: 1340-1-016
CURRENT APPLICATION NUMBER: US/09/073,541A
CURRENT FILING DATE: 1998-05-06
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 25
LENGTH: 8900
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-073-541A-25

Query Match 61.3%; Score 18.4; DB 4; Length 8900;
Best Local Similarity 95.0%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 GTCCCTGGTCATCTCACCCTT 28
|||||
DB 7586 GTCCCGGTGTCATCTCACCCTT 7567

RESULT 15
US-08-961-527-140
Sequence 140, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue

Fri Jul 11 15:29:34 2003

us-10-087-631b-4.rni

Page 6

CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44M storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 140:

SEQUENCE CHARACTERISTICS:

LENGTH: 28882 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-140

Query Match

Best Local Similarity 61.3%; Score 18.4; DB 4; Length 28882;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 2314 GTCTGCGTCATCTACCTT 2333

9 GTCTGCGTCATCTACCTT 28

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

Search completed: July 10, 2003, 20:27:14
Job time : 15.4157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 285.951 Seconds

(without alignments)
3358.597 Million cell updates/sec

Title: US-10-087-631B-5

Perfect score: 33

Sequence: 1 gccacatgagtcgcaagcgctgtgtataccg 33

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : GenEmbl:

1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ov:*
8: gb_ov:*
9: gb_ov:*
10: gb_ov:*
11: gb_ov:*
12: gb_ov:*
13: gb_ov:*
14: gb_ov:*
15: gb_ov:*
16: gb_ov:*
17: gb_ov:*
18: gb_ov:*
19: gb_ov:*
20: gb_ov:*
21: gb_ov:*
22: gb_ov:*
23: gb_ov:*
24: gb_ov:*
25: gb_ov:*
26: gb_ov:*
27: gb_ov:*
28: gb_ov:*
29: gb_ov:*
30: gb_ov:*
31: gb_ov:*
32: gb_ov:*
33: gb_ov:*
34: gb_ov:*
35: gb_ov:*
36: gb_ov:*
37: gb_ov:*
38: gb_ov:*
39: gb_ov:*
40: gb_ov:*
41: gb_ov:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	69.7	84126	9	AL008728 Human DNA
2	23	69.7	156998	2	AL136099
3	21	63.6	178874	2	AC096936 Rattus no
4	20.8	63.0	1333	10	RNO251687
5	20.8	63.0	187614	2	AC104398 Rattus no
6	20.8	63.0	189192	2	AC119125 Rattus no
7	20.8	63.0	205020	2	AC109903 Rattus no
8	20.8	63.0	318696	2	AC102955 Rattus no
9	20.4	61.8	3079	8	D89010
10	20.2	61.2	24263	3	DMC1183
11	20.2	61.2	40396	2	AC087531
12	20.2	61.2	102227	2	AC020129
13	20.2	61.2	106265	9	AL133394
14	20.2	61.2	161668	3	AC104602
15	20.2	61.2	179752	2	AC087522
16	20.2	61.2	188448	2	AC021368
17	20.2	61.2	203322	2	AC115149
18	20.2	61.2	299970	3	AE003418
19	20.2	60.6	696	14	E30279196
20	20	60.6	744	14	E30279190
21	20	60.6	744	14	E30279191
22	20	60.6	744	14	E30279192
23	20	60.6	3185	14	AF179613
24	20	60.6	6489	14	HEC295172
25	20	60.6	80755	2	AC118767
26	20	60.6	149967	2	AC025397
27	20	60.6	158138	2	AC129879
28	20	60.6	161355	9	HSDJ60019
29	19.8	60.0	771	6	AX057978
30	19.8	60.0	70113	2	AC101633
31	19.8	60.0	70113	2	AC101633
32	19.8	60.0	78132	2	AC022794
33	19.8	60.0	78132	2	AC022794
34	19.8	60.0	90150	9	AC068576
35	19.8	60.0	140757	9	AL157778
36	19.8	60.0	154053	2	AC110331
37	19.8	60.0	159507	9	AC004985
38	19.8	60.0	164165	2	AC103385
39	19.8	60.0	165413	9	AC078784
40	19.8	60.0	174327	2	AC068159
41	19.8	60.0	181561	9	AC015911
42	19.8	60.0	202308	2	AL772301
43	19.8	60.0	208752	2	AC124687
44	19.8	60.0	223088	2	AC099383
45	19.8	60.0	239471	2	AC099284

ALIGNMENTS

RESULT 1
HS125N5
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

HS125N5
Human DNA sequence from clone 125N5 on chromosome 6q26-27. Contains a putative novel gene, ESTs, STSS and GSSs, complete sequence.
AL008728
AL008728.1 GI:3820992
HTG.
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 84126)
Mashreghi-Mohammadi, M.
Direct Submission

JOURNAL
Submitted (04-NOV-1998) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Nov 2, 1998 this sequence version replaced gi:3717966.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence is the entire insert of clone 125N5. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>
125N5 is from the library RPL1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:PCRPAC2>

FEATURES

SOURCE

```

1. 84126
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="q26-27"
/chromosome="6"
/clone="RPL-125N5"
/location="1p11-RPCL-1"

68. 374
/note="AluSx repeat: matches 1. 303 of consensus"
1401. 1665
/note="5 copies 53 mer 89% conserved"
1642. 1770
/note="3 copies 43 mer 81% conserved"
join(2407. 2474,3240. 3472)
/gene="dJ125N5.1"
join(<2407. 2474,3240. 3472)
/gene="dJ125N5.1"
/product="dJ125N5.1 (PUTATIVE novel protein)"
/note="match: EST AA732827"
/evidence="not_experimental"
2684. 2794
/note="AluSg/X repeat: matches 178. 288 of consensus"
complement(4097. 4389)
/note="match: GSS A0053978"
4980. 5072
/note="MIR repeat: matches 75. 174 of consensus"
5198. 5306
/note="L2 repeat: matches 21. 136 of consensus"
5628. 5762
/note="L2 repeat: matches 554. 689 of consensus"
7178. 7642
/note="LTR repeat: matches 1. 450 of consensus"
8264. 8481
/note="HERVH repeat: matches 4314. 4531 of consensus"
8575. 8676
/note="HERVH repeat: matches 4170. 4272 of consensus"
8679. 9122
/note="HERVH repeat: matches 3186. 3637 of consensus"
9105. 9287
/note="HERVH repeat: matches 2669. 2847 of consensus"
9293. 9405
/note="MERAD repeat: matches 1. 114 of consensus"
9408. 10330
/note="L1PB1 repeat: matches 5221. 6118 of consensus"
10336. 10415
/note="8 copies 10 mer aaaaagaag 69% conserved"
10340. 10411
/note="36 copies 2 mer aa 68% conserved"

```

```

repeat_region 10467. 10770
/note="AluSg repeat: matches 1. 301 of consensus"
repeat_region 10771. 11083
/note="MER4B repeat: matches 206. 484 of consensus"
repeat_region 11178. 11309
/note="MER4D repeat: matches 865. 989 of consensus"
repeat_region 11809. 12511
/note="MER4D repeat: matches 236. 973 of consensus"
misc_feature complement(12331. 12682)
/note="match: GSS B30601"
12691. 13063
/note="match: GSS B51137"
/note="match(12868. 13508)
/note="match: GSS A0200723"
14553. 14613
/note="LTR16C repeat: matches 308. 373 of consensus"
repeat_region 14650. 14763
/note="L2 repeat: matches 2601. 2709 of consensus"
14764. 15109
/note="match: GSS A0032080"
14881. 14922
/note="MER5A repeat: matches 112. 157 of consensus"
15110. 15197
/note="2 copies 44 mer 86% conserved"
15233. 15576
/note="LTRFB repeat: matches 1. 366 of consensus"
16556. 16713
/note="2 copies 79 mer 82% conserved"
16733. 16843
/note="3 copies 37 mer 83% conserved"
19332. 19642
/note="AluSg repeat: matches 1. 309 of consensus"
19661. 19752
/note="L1M4 repeat: matches 2964. 3055 of consensus"
19753. 20067
/note="AluSg repeat: matches 1. 312 of consensus"
20068. 20298
/note="L1M4 repeat: matches 2734. 2964 of consensus"
20368. 20485
/note="L1M4 repeat: matches 6249. 7977 of consensus"
20571. 20721
/note="MER57-internal repeat: matches 7145. 7305 of consensus"
20848. 20895
/note="MER57-internal repeat: matches 7387. 7434 of consensus"
20932. 21206
/note="L1PAB repeat: matches 5889. 6163 of consensus"
21493. 21583
/note="L1M4 repeat: matches 7492. 7586 of consensus"
21860. 21962
/note="L1M4 repeat: matches 7104. 7205 of consensus"
22229. 22402
/note="LTR1C repeat: matches 285. 461 of consensus"
23093. 23196
/note="HAL1 repeat: matches 680. 781 of consensus"
23650. 23875
/note="L1M6C repeat: matches 2385. 2295 of consensus"
23949. 24038
/note="LTR1C repeat: matches 1. 84 of consensus"
26324. 26634
/note="Alu repeat: matches 1. 311 of consensus"
26811. 26895
/note="L1M2 repeat: matches 6038. 6122 of consensus"
complement(26912. 27500)
/note="match: genomic DNA Z77936"
27022. 27116
/note="L1M2 repeat: matches 6064. 6159 of consensus"
27124. 27180
/note="3 copies 19 mer 91% conserved"
27531. 27817
/note="AluSg repeat: matches 6. 296 of consensus"
28174. 28315

```

```

repeat_region /note="MIR repeat: matches 44. 184 of consensus"
30062. 30361 /note="AlusX repeat: matches 1. 312 of consensus"
repeat_region 30864. 30935
/note="4 copies 18 mer 81% conserved"
repeat_region 31965. 32052
/note="2 copies 44 mer 86% conserved"
repeat_region 32326. 32382
/note="Charlief3 repeat: matches 2644. 2700 of consensus"
repeat_region 33415. 33452
/note="MER63 repeat: matches 3. 38 of consensus"
misc_feature 33890. 34267
/note="match: GSS AQ178363"
repeat_region 34268. 34339
/note="2 copies 36 mer 94% conserved"
repeat_region 35047. 35550
/note="14 copies 36 mer 94% conserved"
repeat_region 35922. 36057
/note="MIR repeat: matches 39. 184 of consensus"
repeat_region 36297. 36364
/note="2 copies 34 mer 96% conserved"
repeat_region 37108. 37258
/note="L1P44 repeat: matches 6000. 6146 of consensus"
repeat_region 38408. 38501
/note="MIR repeat: matches 162. 258 of consensus"
prim_transcript complement(40975. 41312)
/note="match: EST AA883825"
repeat_region 41592. 41711
/note="L2 repeat: matches 2290. 2418 of consensus"
prim_transcript complement(43225. 43547)
/note="match: EST AA977687"
repeat_region 44041. 44093
/note="MIR repeat: matches 59. 113 of consensus"
repeat_region 44143. 44183
/note="MIR repeat: matches 218. 258 of consensus"
repeat_region 44214. 45044
/note="277 copies 3 mer tga 63% conserved"
repeat_region 44214. 45041
/note="46 copies 18 mer 63% conserved"
repeat_region 44215. 45042
/note="23 copies 36 mer 63% conserved"

Query Match 69.7%: Score 23; DB 9; Length 84126;
Best Local Similarity 83.9%: Pred. No. 10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCCACATGAGTGCAGGCGCTGTGTATAC 31
Db 11528 GCCACCTGAGTGCAGGCGCTGTGTGGAAC 11558

RESULT 2
LOCUS AL136099 156998 bp DNA linear HTG 16-AUG-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-270C4, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL136099
VERSION AL136099.10 GI:9796029
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 156998)
AUTHORS Phillimore,B.
TITLE Direct Submission
JOURNAL Submitted (13-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Aug 11, 2000 this sequence version replaced gi:9588114.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC

```

```

Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba270C4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 15572 bases at least Q40
Consensus quality: 15572 bases at least Q40
Consensus quality: 15572 bases at least Q20
Insert size: 156998; sum-of-contigs
Insert size: 157873; 2.1% error; agarose-fp
Quality coverage: 4.85x in Q20 bases; agarose-fp
coverage: 4.85x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
location/Qualifiers
source 1. 156998
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-270C4"
/clone_id="RPOI-11.1"
/note="assembly_fragment:00254
clone_end:77
vector_side:right"

BASE COUNT 42309 a 35053 c 36018 g 43618 t
ORIGIN

Query Match 69.7%: Score 23; DB 2; Length 156998;
Best Local Similarity 83.9%: Pred. No. 10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCCACATGAGTGCAGGCGCTGTGTATAC 31
Db 42942 GCCACCTGAGTGCAGGCGCTGTGTGGAAC 42972

RESULT 3
LOCUS AC096936/C 178874 bp DNA linear HTG 13-JUL-2002
DEFINITION Rattus norvegicus clone CH230-41H3, *** SEQUENCING IN PROGRESS ***,
63 unordered pieces.
ACCESSION AC096936
VERSION AC096936.4 GI:21735373
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 178874)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,W., Banks,T.,
Barbarta,J., Benton,J., Blinage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowle,S., Bileva,M., Brown,M., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Honsi, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kritovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisgied, H.,
Lodado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Madeshwar, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabadi, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Nickerson, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokwenkwo, S., Ogulu, M., Okunnu, G.,
Orduyue, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojudoan, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherrer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, T.,
Soderren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Taylor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Velitod, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, J., Vera, V., Villalón, D., Vinson, R., Wang, O.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

Unpublished
Direct Submission
2 (bases 1 to 178874)

Worley, K.C.
Direct Submission

Submitted (03-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 178874)

Worley, K.C.
Direct Submission

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:17973294.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project information

Center project name: CH230-41H3

Center clone name: CH230-41H3

Sequencing vector: plasmid

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 113609 bases at least Q40

Consensus quality: 119045 bases at least Q30

Consensus quality: 124009 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 63 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1168: contig of 1168 bp in length
1169 1268: gap of unknown length
1269 2835: contig of 1567 bp in length
2836 2935: gap of unknown length
2936 4058: contig of 1123 bp in length
4059 4158: gap of unknown length
4159 5514: contig of 1356 bp in length
5515 5614: gap of unknown length

5615 6702: contig of 1088 bp in length
6703 6703: gap of unknown length
6703 8100: contig of 1307 bp in length
8100 8209: gap of unknown length
8209 9708: contig of 1499 bp in length
9708 9709: gap of unknown length
9709 9808: contig of 1080 bp in length
9808 10888: gap of unknown length
10888 10989: gap of unknown length
10989 12002: gap of unknown length
12002 12002: contig of 1014 bp in length
12002 13623: contig of 1521 bp in length
13623 13724: gap of unknown length
13724 15455: contig of 1732 bp in length
15455 17380: gap of unknown length
17380 17480: contig of 1825 bp in length
17480 19205: gap of unknown length
19205 19306: contig of 1725 bp in length
19306 20801: gap of unknown length
20801 20902: contig of 1496 bp in length
20902 22078: gap of unknown length
22078 22178: contig of 1176 bp in length
22178 23587: gap of unknown length
23587 23688: contig of 1410 bp in length
23688 25661: gap of unknown length
25661 25761: contig of 1974 bp in length
25761 26937: gap of unknown length
26937 27038: contig of 1176 bp in length
27038 28299: gap of unknown length
28299 28300: contig of 1262 bp in length
28300 29950: gap of unknown length
29950 30050: gap of unknown length
30050 31787: contig of 1551 bp in length
31787 31886: gap of unknown length
31886 32898: contig of 1736 bp in length
32898 32998: gap of unknown length
32998 34776: contig of 1012 bp in length
34776 34877: gap of unknown length
34877 36793: contig of 1778 bp in length
36793 36892: gap of unknown length
36892 39218: contig of 1916 bp in length
39218 39219: gap of unknown length
39219 40269: contig of 2226 bp in length
40269 40368: gap of unknown length
40368 42142: contig of 1050 bp in length
42142 42242: gap of unknown length
42242 43554: contig of 1774 bp in length
43554 43654: gap of unknown length
43654 44781: contig of 1312 bp in length
44781 44880: gap of unknown length
44880 47130: contig of 1126 bp in length
47130 47230: gap of unknown length
47230 49647: contig of 2250 bp in length
49647 49747: gap of unknown length
49747 50976: contig of 2417 bp in length
50976 51076: gap of unknown length
51076 53855: contig of 1228 bp in length
53855 53955: gap of unknown length
53955 56999: contig of 2780 bp in length
56999 57098: gap of unknown length
57098 60136: contig of 3043 bp in length
60136 60237: gap of unknown length
60237 62081: contig of 3038 bp in length
62081 62181: gap of unknown length
62181 64959: contig of 1845 bp in length
64959 65059: gap of unknown length
65059 67195: contig of 2778 bp in length
67195 67295: gap of unknown length
67295 69539: contig of 2136 bp in length
69539 69640: gap of unknown length
69640 72537: contig of 2244 bp in length
72537 72637: gap of unknown length
72637 75265: contig of 2898 bp in length
75265 75638: gap of unknown length
75638 75638: contig of 2628 bp in length

```

      /product="pepsinogen F protein"

```

protein_id="CA675982.1"
/db_xref="GI:7105998"
/translation="MKMLVYLGLVALSECLVKRIPKLIKSRMRLEIRSHMILKDYLEYK
PSRRHARLLEQHRNRPNSVTYEPKRNVIIDLIVITGISGTGPQERKRVVLDTSSDILMWSPS
IYCSPACCHAKHPNVFNLQSTELPVSGRPNVAAGSEMSPFLAVTVKIGDLTVVAQA
FLSLSEPEPRFMHEAVFDGILIGLYPNLGOGTVPVDINIMIGLLIQNFALFLSKR
DKGSJLMVGADPVSYHGELHMVPVSKSPWOLAYDSIMNEEIIICDGCGOINDTD
GRSVTGPPSSLTINTONILGARSGSGEVELFKDITLPIDIFTIDSVTPYPASAVY
IKRDHSNRKSNRPFESTDDPSDEPLWLIVGDVFLRLIFTYVDRAANRKIGLASAA"

pclide
pcptide
306 a 361 c 343 g 323 t

/Product="pepsinogen F protein"

ch 63.0%; Score 20.8; DB 10; Length 133;
1 Similarity 78.1%; Pred. No. 1.le+02;
25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

2 CCACATGATGGCAGCAGCGCTTCGTGATACC 33
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CCGCATGGTGCACAGCGTTCTTGATGCCG 930

AC104338 187614 bp DNA linear HTG 12-JUL-2002
Rattus norvegicus clone CH230-128124, *** SEQUENCING IN PROGRESS
AC104338
HTG: HTGS_PHASE1.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 187614)
Muzyr,D.M., Adams,C., Adio-oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Ate,U.R., Ayele,M., Banks,T.,
Barbaris,J., Benton,J., Bimege,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowles,B., Brieva,M., Brown,E., Browm,M., Bryant,N.P.,
Buhyr,C., Burgh,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Delaney,L.M., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Deville,K.R., Delgado,O., Benn,A.-L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Edwards,D., Eugen-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Fallis,T., Ferraguto,D., Flaggs,N., Ford,J., Foster,P., Frantz,P.,
Gabrisl,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Gueraña,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homai,F., Howard,S., Huber,V., Huily,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivert,S., Judah,S.,
Karissom,B., Kelly,S., Khan,U., King,L., Koryak,S., Kovac,C.,
Kraticovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W., Loulsaged,H.,
Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapa,P., Martin,R., Matindale,A., Martinez,E.,
Massey,E., Maxwell,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Mosser/M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nockenkno,S., Oguh,M., Okunolu,G.,
Oragunye,N., Oriedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Petters,L., Plickens,R., Primus,E., Pu,L.L., Qules,M., Ren,X.,
Rivers,M., Rojas,A., Rojoudokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherrer,S., Scott,G., Shen,H., Shoshitar,N., Sisson,J.,
Soedrgren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabori,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,

TITLE
JOURNAL
AUTHORS
REFERENCE
TITLE
JOURNAL
COMMENT

Usumani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczek, K., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. P., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 187614)
Worley, K.C.
Submitted (09-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 187614)
Worley, K.C.
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:18846151.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: G10D
Center clone name: CH230-128124
Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 115292 bases at least Q40
Consensus quality: 120980 bases at least Q30
Consensus quality: 125749 bases at least Q20

NOTE: Estimated insert size may differ from sequence length.
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently consists of 72 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1
1025 1024: contig of 1024 bp in length
1125 1124: gap of unknown length
2125 2124: contig of 1000 bp in length
2225 2224: gap of unknown length
3533 3533: contig of 1309 bp in length
3634 3633: gap of unknown length
4847 4846: contig of 1213 bp in length
4947 4946: gap of unknown length
6171 6170: contig of 1224 bp in length
6271 6270: gap of unknown length
7309 7308: contig of 1038 bp in length
7409 7408: gap of unknown length
8615 8614: contig of 1206 bp in length
8715 8714: gap of unknown length
10067 10066: contig of 1352 bp in length
10167 10166: gap of unknown length
11507 11506: contig of 1340 bp in length
11607 11606: gap of unknown length
12925 12924: contig of 1318 bp in length
13025 13024: gap of unknown length
14558 14557: contig of 1533 bp in length
14658 14657: gap of unknown length
16006 16006: contig of 1349 bp in length
16007 16006: gap of unknown length
17604 17604: contig of 1498 bp in length
17705 17704: gap of unknown length
19136 19136: contig of 1432 bp in length

19137 19236: gap of unknown length
19237 20375: contig of 1139 bp in length
20376 20475: gap of unknown length
20476 21637: contig of 1162 bp in length
21638 21737: gap of unknown length
21738 23001: contig of 1264 bp in length
23002 23101: gap of unknown length
23102 24450: contig of 1349 bp in length
24451 24550: gap of unknown length
24551 26374: contig of 1724 bp in length
26375 27484: contig of 1110 bp in length
27485 27584: gap of unknown length
27585 29207: contig of 1623 bp in length
29208 29308: gap of unknown length
29309 30997: contig of 1689 bp in length
31097 32229: contig of 1133 bp in length
32230 33329: gap of unknown length
33330 34203: contig of 1874 bp in length
34204 34303: gap of unknown length
34304 35693: contig of 1390 bp in length
35694 35793: gap of unknown length
35794 36975: contig of 1182 bp in length
36976 37075: gap of unknown length
37076 38265: contig of 1190 bp in length
38266 38365: gap of unknown length
38366 39489: contig of 1124 bp in length
39490 39589: gap of unknown length
39590 40625: contig of 1036 bp in length
40626 40725: gap of unknown length
40726 41981: contig of 1266 bp in length
41982 42091: gap of unknown length
42092 43128: contig of 1037 bp in length
43129 44328: gap of unknown length
44329 44509: contig of 1281 bp in length
44510 44610: gap of unknown length
44611 46095: contig of 1486 bp in length
46096 46195: gap of unknown length
46196 48271: contig of 2076 bp in length
48272 48371: gap of unknown length
48372 50278: contig of 1907 bp in length
50279 50378: gap of unknown length
50379 53547: contig of 3169 bp in length
53548 53647: gap of unknown length
53648 55387: contig of 1740 bp in length
55388 55487: gap of unknown length
55488 56721: contig of 1234 bp in length
56722 56821: gap of unknown length
56822 59031: contig of 2210 bp in length
59032 59131: gap of unknown length
59132 60975: contig of 1844 bp in length
60976 61075: gap of unknown length
61076 62993: contig of 1918 bp in length
62994 63093: gap of unknown length
63094 64432: contig of 1339 bp in length
64433 64533: gap of unknown length
64533 66619: contig of 2087 bp in length
66620 66719: gap of unknown length
66720 68794: contig of 2075 bp in length
68795 68895: gap of unknown length
68896 71194: gap of unknown length
71195 71294: contig of 2300 bp in length
71295 73659: contig of 2375 bp in length
73660 73769: gap of unknown length
73770 76103: contig of 2334 bp in length
76104 76203: gap of unknown length
76204 78867: contig of 2664 bp in length
78868 78968: gap of unknown length
78969 80169: contig of 1202 bp in length
80170 80269: gap of unknown length
80270 81945: contig of 1676 bp in length
81946 82045: gap of unknown length


```

26436 28199: contig of 1764 bp in length
28200 28299: gap of unknown length
28300 29444: contig of 1145 bp in length
29445 29545: gap of unknown length
29545 31447: contig of 1903 bp in length
31448 31547: gap of unknown length
31548 34038: contig of 2491 bp in length
34039 34138: gap of unknown length
34139 35811: contig of 1673 bp in length
35812 35912: gap of unknown length
35912 38254: contig of 2243 bp in length
38255 38355: gap of unknown length
38355 39887: contig of 1532 bp in length
39887 39986: gap of unknown length
41723 41822: contig of 1736 bp in length
41823 43529: contig of 1706 bp in length
43529 43628: gap of unknown length
43629 45290: contig of 1662 bp in length
45291 45390: gap of unknown length
45391 47399: contig of 2009 bp in length
47400 47500: gap of unknown length
47500 49019: contig of 1519 bp in length
49019 49118: gap of unknown length
49119 51076: contig of 1958 bp in length
51077 51176: gap of unknown length
51177 53555: contig of 2379 bp in length
53556 53655: gap of unknown length
53656 56041: contig of 2386 bp in length
56042 56141: gap of unknown length
56142 58492: contig of 2351 bp in length
58493 58592: gap of unknown length
58593 60409: contig of 1817 bp in length
60410 60510: gap of unknown length
60510 63422: contig of 2913 bp in length
63423 63522: gap of unknown length
63523 63780: contig of 2258 bp in length
63781 65880: gap of unknown length
65881 68421: contig of 2541 bp in length
68422 68522: gap of unknown length
68522 71408: contig of 2887 bp in length
71409 71508: gap of unknown length
71509 74335: contig of 2827 bp in length
74336 74435: gap of unknown length
74436 76581: contig of 2146 bp in length
76582 76681: gap of unknown length
76682 79512: contig of 2831 bp in length
79513 82209: gap of unknown length
82210 82309: contig of 2597 bp in length
82310 86599: gap of unknown length
86600 86699: contig of 4290 bp in length
86700 90718: gap of unknown length
90719 90818: gap of unknown length
90819 94040: contig of 3222 bp in length
94041 94140: gap of unknown length
94141 97476: contig of 3336 bp in length
97477 97576: gap of unknown length
97577 101462: contig of 3886 bp in length
101463 101562: gap of unknown length
101563 105925: contig of 4363 bp in length
105926 106025: gap of unknown length
106026 110657: contig of 4632 bp in length
110658 110757: gap of unknown length
110758 116260: contig of 5503 bp in length
116261 116360: gap of unknown length
116361 120661: contig of 4301 bp in length
120662 125559: gap of unknown length
125560 125659: contig of 4798 bp in length
125660 129795: gap of unknown length
129796 129895: contig of 4136 bp in length
129896 135463: gap of unknown length
135463 135463: contig of 5568 bp in length

```

```

Query Match      63.0%; Score 20.8; DB 2; Length 189192;
Best Local Similarity 78.1%; Pred. No. 1,1e+02;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 2 CCACATGAGTGCACAGCGCTGTGGATACCG 33
DB 30724 CCTCATGTGTGCCAAGCGCTGTGTATGCGC 30693

```

```

RESULT 7
AC109903 205020 bp. DNA Linear HTG 13-JUL-2002
LOCUS Rattus norvegicus clone CH230-115M11, *** SEQUENCING IN PROGRESS
DEFINITION *** 67 unordered pieces.
ACCESSION AC109903.3 GI:21738286
VERSION HTG, HTGS, PHASE1.
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

REFERENCE

AUTHORS

1 (bases 1 to 205020)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,

Bouck,J., Bowler,S., Brileva,M., Brown,M., Brown,N., Brown,D.,

Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,

Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

Devilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,

DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinu,H.H.,

Doutwalte,K.J., Draper,H., Dugan-Rocha,S., Dublin,K.J.,

Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,

Gabriel,J.H., Garcia,A., Garner,T., Garza,N., Gill,R.,

Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,

Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,

Homsl,F., Howard,S., Huber,J., Hulys,S., Hume,J., Jackson,L.E.,

Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudh,S.,

Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,

Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,

Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,

Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,

Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,

Miney,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,

Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,

Nguyen,N., Nickerson,E., Nwokkenko,S., Ogutu,M., Okwoudu,G.,

Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,

Peters,L., Pickens,R., Primus,E., Pu,L.L., Qutles,M., Ren,Y.,

Rivers,M., Rojas,A., Rojoubokan,I., Rolfe,M., Ruiz,S., Savery,G.,

Scherrer,S., Scott,G., Shen,H., Shoochari,N., Sisson,I.,

Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,

Sutton,A., Svatek,A., Taber,P., Tamezisa,A., Tamezisa,K., Tang,H.,

Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,

Umanu,K., Vasquez,L., Vera,V., Villalón,D., Vihson,R., Wang,Q.,

Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,

Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,

Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

Weinstock,G. and Gibbs,R.

TITLE

Unpublished

2 (bases 1 to 205020)

Direct Submission

Worley,K.C.

REFERENCE

Submitted (08-FEB-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 205020)

31554 32635: contig of 1082 bp in length
 32636 32735: gap of unknown length
 32737 33742: contig of 1007 bp in length
 33743 33842: gap of unknown length
 33843 35401: contig of 1559 bp in length
 35402 35501: gap of unknown length
 35502 36514: contig of 1013 bp in length
 36515 36614: gap of unknown length
 36615 37748: contig of 1134 bp in length
 37749 37848: gap of unknown length
 37849 39053: contig of 1205 bp in length
 39054 39153: gap of unknown length
 39154 40437: contig of 1284 bp in length
 40438 40537: gap of unknown length
 40538 41998: contig of 1461 bp in length
 41999 42098: gap of unknown length
 42099 43458: contig of 1360 bp in length
 43459 43558: gap of unknown length
 43559 44573: contig of 1015 bp in length
 44574 44673: gap of unknown length
 44674 45731: contig of 1058 bp in length
 45732 45831: gap of unknown length
 45832 47490: contig of 1659 bp in length
 47491 47591: gap of unknown length
 47592 48619: contig of 1029 bp in length
 48620 48719: gap of unknown length
 48720 49748: contig of 1029 bp in length
 49749 49848: gap of unknown length
 49849 51127: contig of 1279 bp in length
 51128 51227: gap of unknown length
 51228 52354: contig of 1127 bp in length
 52355 52454: gap of unknown length
 52455 53733: contig of 1279 bp in length
 53734 53833: gap of unknown length
 53834 54939: contig of 1106 bp in length
 54940 55039: gap of unknown length
 55041 56467: contig of 1428 bp in length
 56468 56567: gap of unknown length
 56568 57628: contig of 1061 bp in length
 57629 57728: gap of unknown length
 57729 58837: contig of 1109 bp in length
 58838 58937: gap of unknown length
 58939 59946: contig of 1009 bp in length
 59947 60046: gap of unknown length
 60047 61049: contig of 1003 bp in length
 61050 61149: gap of unknown length
 61150 62708: contig of 1359 bp in length
 62709 62808: gap of unknown length
 62809 64277: contig of 1469 bp in length
 64278 64377: gap of unknown length
 64379 66150: contig of 1773 bp in length
 66151 66250: gap of unknown length
 66251 67433: contig of 1183 bp in length
 67434 67533: gap of unknown length
 67534 69376: contig of 1843 bp in length
 69377 69476: gap of unknown length
 69477 70831: contig of 1355 bp in length
 70832 70931: gap of unknown length
 70932 72325: contig of 1394 bp in length

Query Match 63.0%: Score 20.8; DB 2: Length 318696;
 Best Local Similarity 78.1%; Pred. No. 1.1e+02;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 CCACATGATGGCAGCGCTGTGATACCG 33
 Db 298103 CCTCATGTGTCGCAAGCGCTGTGATACCG 298134

RESULT 9
 LOCUS D89010 3079 bp DNA linear PLN 12-MAY-1999
 DEFINITION Aspergillus oryzae DNA for BRLA, complete cds.

ACCESSION D89010 GI:1694664
 VERSION D89010.1
 KEYWORDS BRLA.
 SOURCE Aspergillus oryzae DNA.
 ORGANISM Aspergillus oryzae
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

REFERENCE
 AUTHORS Yamada, O., Lee, R. R., Gomi, K. and Iimura, Y.
 TITLE Cloning and functional analysis of the Aspergillus oryzae
 conidiation regulator gene brla by its disruption and misscheduled
 expression
 JOURNAL J. Biosci. Bioeng. 87, 424-429 (1999)
 REFERENCE 2 (bases 1 to 3079)
 AUTHORS Yamada, O.
 TITLE Direct Submission
 JOURNAL Submitted (12-NOV-1996) Osamu Yamada, National Research Institute
 of Brewing, Microbiology Division; 3-7-1 Kagamiyama,
 Hiasahi-Hiroshima, Hiroshima 739, Japan
 (E-mail: Yamada.Oenr@ib.go.jp, Tel: 81-824-20-0824,
 Fax: 81-824-20-0808)

FEATURES
 source
 1..3079
 Location/Qualifiers
 /organism="Aspergillus oryzae"
 /db_xref="taxon:5062"
 1449..2714
 /gene="brla"
 1449..2714
 /gene="brla"
 /codon_start=1
 /product="BRLA"
 /protein_id="BAA13744.1"
 /db_xref="GI:1694665"
 /translation="MRTQNNLTVEVDCHSLGNSNCPSTSSFSFSDSPTPTPTISQ
 GSLASPGMODAGSYGHAVERHTGATPRSAFRLAGTNSNMGMSTGAMAOERMP
 PDFLSAYDNDYHEWMLPSDGRKYEYGLPYHTLPQCPTNRYNVPRAALPEA
 ATNCLSRISFIHAEVPOSMGMNMPWIOAESIAPORTIASOVGVPVPPSYSE
 EPFSTIOFKHTSPPTPLRSCISGASGDPDPSLSSGAADYLDFOOSPFRGLNR
 LORPDSRMTRKSSRONMSEINPSYIKKOPFCRKCRCGRKRDHKKRHHKSL
 KERHVCNVECCERAFSRSDNLNHYTHSKRGGRNRYVATLDESSFDYDPRGOL
 TPDELPTKGLSTLDDMPNRSRYSDGLDD"

BASE COUNT 728 a 858 c 587 g 906 t
 ORIGIN

Query Match 61.8%: Score 20.4; DB 8: Length 3079;
 Best Local Similarity 80.0%; Pred. No. 1.7e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 3 CACATGATGGCAGCGCTGTGATACCG 32
 Db 1594 CACCTGGGTGCAAGACCGCGGTCAATACC 1623

RESULT 10
 LOCUS DMC118B3/C 24263 bp DNA linear INV 27-APR-1999
 DEFINITION Drosophila melanogaster cosmid clone 118B3.
 ACCESSION AL031582
 VERSION AL031582.1 GI:4176433
 KEYWORDS
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 24263)
 AUTHORS Philippe, Valenti., Catherine, Salles., Lorna, Campbell. and
 David, Glover.
 TITLE Sequencing the distal x chromosome of Drosophila melanogaster
 JOURNAL Unpublished
 DEPARTMENT Department of Anatomy and Physiology, Medical Sciences Institute,
 University of Dundee, Dundee DD1 4HN, U.K.
 REFERENCE 2 (bases 1 to 24263)

DEFINITION Homo sapiens chromosome 11 clone RP5-1135K18 map 11, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC087531
VERSION AC087531.3 GI:13431031
KEYWORDS HTG; HTGS-PHASEO.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 40396)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE 1 Homo sapiens chromosome 11, clone RP5-1135K18
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 40396)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Batlien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Lakoque,K., Lamazares,R., Landers,T.,
Lencock,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Marquis,N., Mathews,C., McCarthy,M., McEwan,P., McKernan,K.,
Mcneeters,R., Meldrim,J., Menus,L., Mihova,T., Mlenka,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Kogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Sougnaz,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Trivers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 22, 2001 this sequence version replaced gi:13184089.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12143
Center clone name: 1135_K_18

* NOTE: This record contains 62 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1. 570: contig of 570 bp in length
* 571 670: gap of 100 bp
* 671 1229: contig of 559 bp in length
* 1230 1329: gap of 100 bp
* 1330 1887: contig of 558 bp in length
* 1888 1987: gap of 100 bp
* 1988 2527: contig of 540 bp in length
* 2528 2627: gap of 100 bp
* 2628 3176: contig of 549 bp in length
* 3177 3276: gap of 100 bp
* 3277 3829: contig of 553 bp in length

3830 3929: gap of 100 bp
* 3930 4492: contig of 563 bp in length
* 4493 4592: gap of 100 bp
* 4593 5148: contig of 556 bp in length
* 5149 5248: gap of 100 bp
* 5249 5781: contig of 533 bp in length
* 5782 5881: gap of 100 bp
* 5882 6442: contig of 561 bp in length
* 6443 6542: gap of 100 bp
* 6543 7096: contig of 554 bp in length
* 7097 7197: gap of 100 bp
* 7197 7743: contig of 547 bp in length
* 7744 7843: gap of 100 bp
* 7844 8394: contig of 551 bp in length
* 8395 8494: gap of 100 bp
* 8495 9054: contig of 560 bp in length
* 9055 9154: gap of 100 bp
* 9155 9693: contig of 539 bp in length
* 9694 9793: gap of 100 bp
* 9794 10345: contig of 552 bp in length
* 10346 10445: gap of 100 bp
* 10446 10998: contig of 553 bp in length
* 10999 11098: gap of 100 bp
* 11099 11634: contig of 536 bp in length
* 11635 11734: gap of 100 bp
* 11735 12296: contig of 562 bp in length
* 12297 12396: gap of 100 bp
* 12397 12971: contig of 575 bp in length
* 12972 13071: gap of 100 bp
* 13072 13637: contig of 566 bp in length
* 13638 13737: gap of 100 bp
* 13738 14298: contig of 561 bp in length
* 14299 14398: gap of 100 bp
* 14399 14957: contig of 555 bp in length
* 14958 15057: gap of 100 bp
* 15058 15588: contig of 531 bp in length
* 15589 15688: gap of 100 bp
* 15689 16245: contig of 557 bp in length
* 16246 16345: gap of 100 bp
* 16346 16899: contig of 554 bp in length
* 16900 16999: gap of 100 bp
* 17000 17554: contig of 555 bp in length
* 17555 17654: gap of 100 bp
* 17655 18209: contig of 555 bp in length
* 18210 18309: gap of 100 bp
* 18310 18867: contig of 558 bp in length
* 18868 18967: gap of 100 bp
* 18968 19505: contig of 538 bp in length
* 19506 19605: gap of 100 bp
* 19606 20169: contig of 564 bp in length
* 20170 20269: gap of 100 bp
* 20270 20831: contig of 562 bp in length
* 20832 20931: gap of 100 bp
* 20932 21479: contig of 548 bp in length
* 21480 21579: gap of 100 bp
* 21580 22141: contig of 562 bp in length
* 22142 22241: gap of 100 bp
* 22242 22794: contig of 553 bp in length
* 22795 22894: gap of 100 bp
* 22895 23445: contig of 551 bp in length
* 23446 23545: gap of 100 bp
* 23546 24074: contig of 529 bp in length
* 24075 24174: gap of 100 bp
* 24175 24723: contig of 549 bp in length
* 24724 24823: gap of 100 bp
* 24824 25374: contig of 551 bp in length
* 25375 25474: gap of 100 bp
* 25475 26022: contig of 548 bp in length
* 26023 26122: gap of 100 bp
* 26123 26683: contig of 561 bp in length
* 26684 26783: gap of 100 bp
* 26784 27303: contig of 520 bp in length
* 27304 27403: gap of 100 bp


```

27404 27958: contig of 555 bp in length
27959 28058: gap of 100 bp
28059 28614: contig of 556 bp in length
28615 28714: gap of 100 bp
28715 29270: contig of 556 bp in length
29271 29370: gap of 100 bp
29371 29928: contig of 558 bp in length
29929 30028: gap of 100 bp
30029 30580: contig of 552 bp in length
30581 30680: gap of 100 bp
30681 31229: contig of 549 bp in length
31230 31329: gap of 100 bp
31330 31882: contig of 553 bp in length
31883 31982: gap of 100 bp
31983 32538: contig of 556 bp in length
32539 32638: gap of 100 bp
32639 33187: contig of 549 bp in length
33188 33287: gap of 100 bp
33288 33837: contig of 550 bp in length
33838 33937: gap of 100 bp
33938 34484: contig of 547 bp in length
34485 34584: gap of 100 bp
34585 35145: contig of 561 bp in length
35146 35245: gap of 100 bp
35246 35799: contig of 554 bp in length
35800 35899: gap of 100 bp
35900 36445: contig of 546 bp in length
36446 36545: gap of 100 bp
36546 37103: contig of 558 bp in length
37104 37203: gap of 100 bp
37204 37760: contig of 557 bp in length
37761 37860: gap of 100 bp
37861 38424: contig of 564 bp in length
38425 38524: gap of 100 bp
38525 39071: contig of 547 bp in length
39072 39171: gap of 100 bp
39172 39735: contig of 564 bp in length
39736 39835: gap of 100 bp
39836 40396: contig of 561 bp in length.
Location/Qualifiers
1. 40396
/Organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP5-1135K18"
/clone_11b="RPCI Human PAC library 5"
9313 a 7763 c 7710 g 9183 t 6427 others

Query Match
Best Local Similarity 61.2%; Score 20.2; DB 2; Length 40396;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 GCCACATGAGTGGCAGGCGCTGCGATCCG 33
Db 6231 GCCACAGGAGGGCGAGGAGTGGTGCATCCAG 6199

RESULT 12
AC020129 102227 bp DNA linear HTG 03-JAN-2000
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
ACCESSION AC020129
VERSION AC020129.1 GI:6664768
SOURCE HTG: PHASE2.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 102227)

```

```

AUTHORS Adams M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10212145 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..102227
/Organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 27684 a 23342 c 23493 g 27708 t
ORIGIN
Query Match
Best Local Similarity 61.2%; Score 20.2; DB 2; Length 102227;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 GCCACATGAGTGGCAGGCGCTGCGATCCG 33
Db 91773 GTCGAGGAGTGTGATCGCGCTGCGATCCG 91805

RESULT 13
AL133294/c 106265 bp DNA linear PRI 27-SEP-2000
LOCUS Human DNA sequence from clone RP5-1135K18 on chromosome 11p12-14.1
ACCESSION AL133294
VERSION AL133294.10 GI:9650513
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 106265)
AUTHORS Moore, M.
JOURNAL Direct Submission
COMMENT Submitted (15-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Aug 1, 2000 this sequence version replaced gi:9501134.
Where difference assembly data is compared from overlapping clones,
together with a note of the overlapping clone name. Note that the
corresponding to the overlapping clone. Note that the
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep
RP5-1135K18 is
institute by the group of Pieter de Jong. For further details see
http://bacpac.med.buffalo.edu/
VECTOR: PCYPAC2
IMPORTANT: This sequence is not the entire insert of clone
RP5-1135K18 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP5-1135K18 is at 1 in this sequence.
The true left end of clone RP5-1169J3 is at 106166 in this
sequence.
Location/Qualifiers

```



```
source
1..106265
/organism="Homo sapiens"
/dbxref="taxon:9606"
/chromosome="11"
/map="p12-14.1"
/clone="RP5-1135K18"
/clone.lib="RPC1-5"
1279..1591
repeat_region
/note="AluX repeat: matches 1..311 of consensus"
1592..1675
repeat_region
/note="42 copies 2 mer tt 71% conserved"
1701..1895
repeat_region
/note="L1ME3 repeat: matches 5927..6123 of consensus"
1902..2049
repeat_region
/note="MIR repeat: matches 92..249 of consensus"
2908..3094
repeat_region
/note="L2 repeat: matches 2426..2624 of consensus"
3089..3363
misc_feature
/note="match: GSS: Em:B01382"
complement(3349..3742)
/note="match: GSS: Em:AQ019578"
4577..4790
repeat_region
/note="MIR repeat: matches 48..261 of consensus"
6479..6734
repeat_region
/note="MIR repeat: matches 6..262 of consensus"
7746..7969
repeat_region
/note="L1MEC repeat: matches 2177..2401 of consensus"
7970..8277
repeat_region
/note="AluX repeat: matches 1..303 of consensus"
8278..8419
repeat_region
/note="L1MEC repeat: matches 2377..2177 of consensus"
8420..8724
repeat_region
/note="AluX repeat: matches 1..298 of consensus"
8725..8846
repeat_region
/note="L1MEC repeat: matches 2257..2378 of consensus"
9041..9084
repeat_region
/note="22 copies 2 mer tt 81% conserved"
9087..9396
repeat_region
/note="AluX repeat: matches 1..312 of consensus"
9479..9676
repeat_region
/note="MIR repeat: matches 49..262 of consensus"
9967..10097
repeat_region
/note="Aluub repeat: matches 3..98 of consensus"
complement(10129..10648)
/note="match: GSS: Em:AQ771611"
complement(10161..10661)
/note="match: GSS: Em:AQ353682"
complement(10214..10681)
/note="match: GSS: Em:AQ386199"
10796..10853
repeat_region
/note="MIR repeat: matches 149..207 of consensus"
10855..11158
repeat_region
/note="AluSq repeat: matches 1..308 of consensus"
11251..11515
repeat_region
/note="L2 repeat: matches 2451..2749 of consensus"
11648..11851
repeat_region
/note="L1MC5 repeat: matches 7705..7917 of consensus"
11859..12150
repeat_region
/note="AluX repeat: matches 1..295 of consensus"
12175..12366
repeat_region
/note="Aluub repeat: matches 89..273 of consensus"
12472..12743
repeat_region
/note="AluX repeat: matches 26..301 of consensus"
complement(13148..13905)
/note="match: GSS: Em:B05804"
complement(13300..13785)
/note="match: GSS: Em:AQ432970"
13309..13539
repeat_region
/note="MIR repeat: matches 262..478 of consensus"
complement(13345..13683)
/note="match: GSS: Em:AQ091582"
complement(13384..13771)
/note="match: GSS: Em:AQ005856"

misc_feature
complement(13441..13785)
/note="match: GSS: Em:B50430"
complement(13461..13785)
/note="match: GSS: Em:B78931 Em:B88270"
complement(13829..14267)
/note="match: GSS: Em:AQ228611"
13858..14164
repeat_region
/note="AluSp repeat: matches 1..308 of consensus"
complement(14001..14450)
/note="match: GSS: Em:AQ428027"
complement(14063..14269)
/note="match: GSS: Em:B53280"
15018..15097
repeat_region
/note="AluS repeat: matches 2..82 of consensus"
15098..15402
repeat_region
/note="AluX repeat: matches 1..306 of consensus"
15415..15540
repeat_region
/note="MIR repeat: matches 13..144 of consensus"
15775..16192
misc_feature
/note="match: GSS: Em:AQ084215"
16628..16795
repeat_region
/note="L1PA16 repeat: matches 5823..5965 of consensus"
16796..17090
repeat_region
/note="AluX repeat: matches 1..288 of consensus"
17091..17263
repeat_region
/note="L1PA16 repeat: matches 5965..6146 of consensus"
17286..17574
repeat_region
/note="Aluub repeat: matches 1..300 of consensus"
20176..20349
repeat_region
/note="MER45 repeat: matches 2..178 of consensus"
20396..20571
repeat_region
/note="MIR repeat: matches 64..244 of consensus"
20607..20731
repeat_region
/note="L2 repeat: matches 2627..2749 of consensus"
21239..21369
repeat_region
/note="MIR repeat: matches 116..262 of consensus"
22075..22197
repeat_region
/note="MIR repeat: matches 19..148 of consensus"
22199..22357
repeat_region
/note="Aluub repeat: matches 135..293 of consensus"
22895..23206
repeat_region
/note="AluX repeat: matches 1..312 of consensus"
24362..24506
repeat_region
/note="L2 repeat: matches 2308..2439 of consensus"
24840..24877
repeat_region
/note="L2 repeat: matches 1656..1693 of consensus"
25110..25504
repeat_region
/note="MER69 repeat: matches 1995..2454 of consensus"
25503..25653
repeat_region
/note="MER69 repeat: matches 419..568 of consensus"
25738..26004
repeat_region
/note="MER69 repeat: matches 1..303 of consensus"
26167..26477
repeat_region
/note="AluSq repeat: matches 1..313 of consensus"
26836..27008
repeat_region
/note="MIR repeat: matches 89..262 of consensus"
27070..27363
repeat_region
/note="AluSq repeat: matches 1..295 of consensus"
28049..28082
repeat_region
/note="L2 copies 2 mer tt 94% conserved"
28140..28376
misc_feature
/note="match: GSS: Em:AQ079942"
28309..28488
repeat_region
/note="FRAM repeat: matches 0..176 of consensus"
28521..28828
repeat_region
/note="AluX repeat: matches 3..311 of consensus"
28865..29257
misc_feature
/note="match: GSS: Em:AQ061107"
29126..29270
repeat_region
/note="L1MC4 repeat: matches 7676..7815 of consensus"
29271..29366
repeat_region
/note="L1MC4 repeat: matches 7877..7975 of consensus"
30243..30349
repeat_region
```

	Query March	61.2%	Score 20.2;	DB 9.	Length 106265;
	Best Local Similarity	75.8%	Ped No. 2.2e+02;		
Matches	25; Conservative	0;	Mismatches	8;	Indels 0; Gaps 0;
-0y	1 GCCACATGAGTGCAGAGCGCTGGTATTACCG	33			
Db	92630 GCCACAGAGGAGGCGCAGGAGGTGTGATGTCAG	92598			

RESULT 14	AC104602/c	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
	AC104602		161668 bp DNA linear INV 06-APR-2002	Drosophila melanogaster X BAC R98-1666 (Roswell Park Cancer Institute Drosophila BAC library) complete sequence.	AC104602		GI:18030109	Drosophila melanogaster
								Drosophila melanogaster

Yocum, J. D., Taber, P., Williamson, A., Homi, F. H.,
Bogdan-Rocha, S. D., Sodergren, E. S., Hodgson, A. H., Chen, R. C.,
Ayala, M. A., Scott, G. S., Worley, K. W., Amamatiadis, P. G., Brandon, C.,
Rogers, Y., An, H., Baldwin, D., Beeson, K. Y., Brown, M., Buhay, C.,
Busam, D. A., Center, A., Chen, G., Chen, Z., Clerc-Blankenburg, K.,
Davenport, L. B., Dietz, S. M., Ding, Y., Dodson, K., Doup, L. E.,
Draper, H., Emery-Cohen, A., Ferlita, S., Gang, N. D. S., Houck, J.,
Hostin, D., Howland, T. J., Hume, J., Idegum, C., Jallali, M., Kovar, C.,
Liu, W., Mactel, B., McIntosh, T. C., Morgan, M., Moy, M., Murphy, B.,
Nelson, K. A., Ndassa, Y., Nguyen, N., Perez, L., Pittman, G. S., Puri, V.,
Schaefer, F., Shen, H., Strong, R., Tector, C., Wang, Q., Williams, S. M.,
Xiang, J., Zaveri, J. S., Zhou, J., Zorrilla, S., Smith, H. O.,
Wheeler, D., Weinstein, G., Gibbs, R. and Venter, J. C.

TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS
Direct Submission	Unpublished	2 (bases 1 to 161668)	Worley, K.C.	Direct Submission	Submitted (14-DEC-2001)	Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston TX 77030, USA	Worley, K.C., Adams, C.

REFERENCE
AUTHORS

3 (bases 1 to 161668) 31/03/93, USA

Worley, K.C., Adams, C.C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbroke, S.L., Anaratunga, H.C., Are, J.R., Banks, T., Barbra, J., Beaton, J., Bimege, K., Blankenburg, K., Bonnin, D., Bouck, J., Boyte, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhey, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F., Catter, M., Cavazos, S.R., Chacko, J., Chen, Z., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Deann, A.L., Ding, Y., Dim, H.H., Douthwaite, K.J., Dipter, H., Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flaagg, N., Ford, J., Foster, P., Frenzt, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Homst, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,

TITLE Direct Submission
JOURNAL Submitted (01-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (pages 1 to 16166)
AUTHORS Worley, R.C., Adams, C., Adio-Ogundola, R., Al-Garni, R., Al-

INSUOKOOS, S. L., Amaratingue, H. C., Are, J. R., Banks, T., Barbala, J.,
 Benton, J., Blimage, K., Blankenburg, K., Bonnin, D., Bouck, J.,
 Bowles, S., Brivetta, M., Brown, E., Brown, M., Bryant, N. P., Buhay, C.,
 Burch, P., Burnett, C., Burrell, K. L., Byrd, N. C., Carton, T. F.,
 Career, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
 Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C. D., Cox, C.,
 Coyle, M. D., Dethorne, S. R., David, R., Davila, M. L., Davis, C.,
 Day, Carroll, L., Dederich, D. A., Delaney, K. R., Delgado, O.,
 Dean, A. L., Ding, Y., Dinh, H. H., Douthwaite, K. J., Draper, H.,
 Dugen-Bocha, S., Durbin, K. J., Earnhart, C., Edgar, D., Edwards, C. C.,
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
 Foster, P., Franz, P., Gabisl, A., Gao, J., Garcia, A., Garner, T.,
 Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S.,
 Hamilton, K., Harris, C., Harris, J., Hart, M., Haylak, P., Hayes, A.,
 Hernandez, J., Hernandez, O., Hodson, A., Hogues, M., Holloway, C.,
 Hollins, B., Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
 Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Joliver, S.,
 Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
 Kovar, C., Kratoch, J., Kuresh, S., Kham, U., Ring, L., Koryak, J.,
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
 Lounsgaard, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Marindale, A.,
 Martinez, E., Massey, E., Maynapp, E., McLeod, M. P., Meador, M.,
 Mei, G., Metker, M., Miner, G., Miner, E., Mitchell, T., Mohabhat, K.,
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
 Ogulu, M., Okunolu, G., Oragunla, N., Oviedo, R., Pace, A., Payton, B.,
 Peery, J., Perez, L., Peters, L., Plickens, R., Primus, E., Pu, L.,
 Quiles, M., Ren, Y., Rives, M., Rojes, A., Rojchokan, I., Rolfe, M.,
 Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N.,
 Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
 Stone, H., Sutton, A., Svatek, A., Taylor, C., Taylor, P., Tamerisa, K.,
 Thomas, S., Tanssey, J., Taylor, C., Taylor, P., Tamerisa, K., Thomas, S.,
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Viallato, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wlecyk, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorilla, S., Nelson, D.,
 Weststock, G. and Gibbs, R.

TITLE Direct Submission
JOURNAL Submitted (06-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jan 1, 2002 this sequence version replaced gi:17976408.
INFORMATION: http://www.hgsc.bcm.tmc.edu/or_email_gc-help@bcm.tmc.edu


```
misc_feature 1.664
misc_feature /note="assembly-fragment" 765.1403
misc_feature /note="assembly-fragment" 1504.98418
misc_feature /note="assembly-fragment" 98519.109646
misc_feature /note="assembly-fragment" 109747.117552
misc_feature /note="assembly-fragment" 137663.116971
misc_feature /note="assembly-fragment" 177072.179752
misc_feature
```

BASE COUNT	ORIGIN
46238 a.	43601 c
43184 g	46106 t
623 others	

Query Match	61.2%	Score 20.2	DB 2	Length 179752
Best Local Similarity	75.8%	Pred. No. 2.2e+02		
Matches	25	Conservative	10	Mismatches 8
				Indels 0
				Gaps 0

DY 1 GCCACATGAGTGGCAAGCGCTGTGGTGATACC 33
||||| ||| |||| |..| ||||| ||
Db 65594 GCCACAGGAGGGGCAGGGAGTTGGTGATCCAG 65562

Search completed: July 10, 2003, 19:20:20
Job time : 294.951 secs

C	10	18.6	56.4	973	21	AAC834442	Arabidopsis thaila
C	11	18.6	56.4	1122	22	AA834436	DNA encoding novel
C	12	18.6	56.4	249487	24	ABN85733	Mouse genomic regi
C	13	18.4	55.8	1131	21	AAA96719	Polynucleotide resc
C	14	18.4	55.8	3425	22	AAF93747	Human cDNA encodin
C	15	18.2	55.2	518	21	AAZ48806	Rice Inositol 1,3,
C	16	18.2	55.2	755	22	ABAI1910	Human nervous syst
C	17	18.2	55.2	853	22	AAAT4749	Human reproductive
C	18	18.2	55.2	1740	18	AAAT4749	RHDV capsid gene.
C	19	18.2	55.2	4372	22	ABA20664	Human nervous syst
C	20	18.2	55.2	7785	20	AAV81446	Pig p105 zona pell
C	21	18	54.5	350	23	AA82637	DNA encoding novel
C	22	18	54.5	351	15	AAO63907	Heavy chain variab
C	23	18	54.5	446	23	AAST70318	DNA encoding novel
C	24	18	54.5	500	22	AAK77964	Human immune/hae
C	25	18	54.5	593	22	ABA60417	Human foetal liver
C	26	18	54.5	593	22	AAK08695	Human brain expres
C	27	18	54.5	593	22	AAK34582	Human bone marrow
C	28	18	54.5	593	22	AAI14301	Probe t8987 used t
C	29	18	54.5	593	24	ABS09213	Human genome-deriv
C	30	18	54.5	654	24	ABQ60434	Human Colon cancer
C	31	18	54.5	864	23	AAST70317	DNA encoding novel
C	32	18	54.5	1014	23	AA64291	DNA encoding novel
C	33	18	54.5	1014	23	AAST71765	DNA encoding novel
C	34	18	54.5	1014	23	AA888225	DNA encoding novel
C	35	18	54.5	1100	22	AAFP31252	Human Immune respo
C	36	18	54.5	1140	23	AAAS81729	DNA encoding novel
C	37	18	54.5	1322	23	AAAS81757	DNA encoding novel
C	38	18	54.5	2187	23	AAAS80450	DNA encoding novel
C	39	18	54.5	2541	23	AAAS88228	DNA encoding novel
C	40	18	54.5	2658	23	AAAS81745	DNA encoding novel
C	41	18	54.5	2660	23	AAAS82684	DNA encoding novel
C	42	18	54.5	2901	23	AAST78375	DNA encoding novel
C	43	18	54.5	3663	23	ABL23920	Drosophila melanog
C	44	18	54.5	6261	23	ABL03774	Drosophila melanog
C	45	18	54.5	34634	24	AAD31198	Human WtL-1 genom

ALIGNMENTS

RESULT 1	
ID	AAC86584 standard; DNA; 771 BP.
XX	AAC86584;
AC	
DT	02-APR-2001 (first entry)
XX	
DE	DNA encoding a fusion of a single chain antibody and streptavidin.
KW	Streptavidin; tumour cell; cancer; adenocarcinoma;
KM	hematological malignancy; ss.
XX	
OS	Synthetic.
OS	Streptomycies avidinl1.
OS	Unidentified.
PN	MO200075333-A1.
PD	14-DEC-2000.
PF	05-JUN-2000; 2000MO-US15595.
PR	07-JUN-1999; 99US-0137900.
PR	03-DEC-1999; 99US-0168976.
PA	(NEOR-) NEORX CORP.
PI	Goshorn SC, Graves SS, Schultz JE, Lin Y, Sanderson JA, Reno JM;
DR	WPI; 2001-091213/10.
XX	

PT New vector constructs for expressing genomic streptavidin fusion
PT proteins which are useful for targeting tumour cells associated with
XX cancer, e.g. adenocarcinomas

PS Example 5; Page 93; 100pp; English.

CC The present sequence encodes a fusion of an anti-CEA single chain
CC antibody and streptavidin. The fusion protein is expressed using
CC vectors of the invention. The specification describes vector constructs
CC for expressing streptavidin fusion proteins. The vector comprises a
CC nucleic acid encoding streptavidin or its functional variant operatively
CC linked to a promoter, and a cloning site for insertion of a second
CC streptavidin, interspersed between the promoter and the first nucleic
CC acid sequence. Alternatively, the vector construct comprises a nucleic
CC acid, operatively linked to a promoter, encoding a polypeptide to be
CC fused with streptavidin, and a cloning site for insertion of a second
CC nucleic acid encoding at least 129 amino acids of streptavidin or its
CC functional variant. The fusion proteins are useful for targeting tumour
CC cells, particularly tumour cells associated with cancer.
CC e.g. adenocarcinomas or hematological malignancies. The vector construct
CC is useful for expressing of streptavidin fusion proteins. In particular,
CC these are useful as tools for medical diagnostics and therapeutic
CC purposes, e.g. for detecting the presence or absence of, or treating, a
CC target site within a mammalian host.

SQ Sequence 771 BP; 172 A; 180 C; 230 G; 189 T; 0 other;

Query Match
Best Local Similarity 60.0%; Score 19.8; DB 22; Length 771;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 CCACATGAGTGGCAGGCGCTGCTGATACC 32
DB 54 CAAGATCAGCTGCAAGCGCTTGGTATACC 84

RESULT 2

ID AAA67547/c
AAA67547 standard; DNA; 415 BP.

AC AAA67547;
DT 31-OCT-2000 (first entry)

DE Pinus radiata sucrose phosphate synthase DNA SEQ ID NO:548.

KW Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW transgenic plant; ds.

OS Pinus radiata.

PN WO200022092-A2.

PD 20-APR-2000.

PF 08-OCT-1999; 99WO-NZ00169.

PR 13-OCT-1998; 98US-0170862.

PR 11-AUG-1999; 99US-0148426.

PA (GENE-) GENESIS RES & DEV CORP LTD.
(FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Bloksberg LN;

DR WPI; 2000-339328/29.

PT New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant

PS Claim 1; Page 212-213; 301pp; English.

CC The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an expectation (E) value of 0.01 or less compared to the 835 sequences,
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC given in the present invention.

SQ Sequence 415 BP; 120 A; 85 C; 98 G; 112 T; 0 other;

Query Match
Best Local Similarity 58.8%; Score 19.4; DB 21; Length 415;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 CCACATGAGTGGCAGGCGCTGCTGATATA 30
DB 156 CCATATGGTGGCAGGCGCTTACGACATA 128

RESULT 3

ID AAF15168/c
AAF15168 standard; cDNA; 634 BP.

AC AAF15168;

DT 13-MAR-2001 (first entry)

DE Trichoderma reesei EST SEQ ID NO:7691.

KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;

KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway; engineering; catabolic pathway engineering; ss.

OS Trichoderma reesei.

PN WO200056762-A2.

PD 28-SEP-2000.

PF 22-MAR-2000; 2000WO-US07781.

PR 22-MAR-1999; 99US-0273623.

PA (NOVO) NOVO NORDISK BIOTECH INC.
(NOVO) NOVO NORDISK AS.

XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

DR WPI; 2000-594572/56.

PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags

PS Claim 89; Page 3104; 3161pp; English.

CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring

The present invention describes a method for monitoring differential expression of genes in a first filamentous fungal (FF) cell relative to expression of the same genes in one or more second filamentous fungal cells. The method uses fluorescence-labeled nucleic acids isolated from the FF cells and a substrate of expressed sequence tags (EST). The ESTs are used in the methods for monitoring differential expression of genes

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags

for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.

Sequence 1284 BP; 336 A; 308 C; 329 G; 311 T; 0 other;

Query Match 58.2%; Score 19.2; DB 23; Length 1284; Best Local Similarity 75.0%; Pred. No. 99; Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 2 CCACATGATGCGAAGCGCTGCTGATACCG 33
Db 50 CCACATGCAAGCAAGCAAGCGATGATGATCCG 19

RESULT 6

AA222700 standard; DNA; 5660 BP.

AA222700;

15-FEB-2000 (first entry)

Pseudomonas fluorescens ABC transporter cassette operon I.

ATP-binding cassette; transporter; operon; LipBPD; Serratia marcescens; microbe; protein secretion; ds.

Pseudomonas fluorescens.

Location/Qualifiers

Key CDS 778..2529

product= "component A"

product= "component B"

product= "component C"

JP11276172-A.

12-OCT-1999.

27-MAR-1998; 98JP-0080597

27-MAR-1998; 98JP-0080597

(TANA) TANABE SEIYAKU CO.

WPI; 1999-626936/54.

P-PSDB; AAY55919, AAY55920, AAY55921.

An ABC transporter gene - used to increase the ability of a microbe to secrete proteins

Claim 4; Page 12-14; 28pp; Japanese.

This sequence represents an ATP-binding cassette (ABC) transporter

operon from Pseudomonas fluorescens strain 33. The operon comprises 3 genes where the termination codon of the first gene overlaps with the initiation codon of the second gene. The sequences of the encoded proteins have 60, 44 and 46% homology respectively to the LipBPD proteins from Serratia marcescens. The novel gene and protein can give or increase the ability of a microbe for secreting a protein.

Sequence 5660 BP; 1131 A; 1817 C; 1736 G; 976 T; 0 other;

Query Match 58.2%; Score 19.2; DB 20; Length 5660; Best Local Similarity 75.0%; Pred. No. 1.2e+02; Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 GCCACATGATGCGAAGCGCTGCTGATACCG 32
Db 365 GCCTGACGATGCGAAGCGATGCTGATTTACC 396

RESULT 7

ABN18387/C standard; cDNA; 379 BP.

ABN18387;

24-JUN-2002 (first entry)

Human ORFX polynucleotide sequence SEQ ID NO:5251.

Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis; hyperproliferative disorder; psoriasis; benign tumour; haemorrhage; degenerative disorder; osteoarthritis; neurodegenerative disorder; cardiovascular disease; diabetes mellitus; systemic lupus erythematosus; hypertension; hypothyroidism; cholesterol ester storage disease; immune deficiency; immune disorder; infectious disease; autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis; myasthenia gravis; gene; ss.

Homo sapiens.

NO200192523-A2.

06-DEC-2001.

29-MAY-2001; 2001MO-US10836.

30-MAY-2000; 2000US-206132P.

29-AUG-2000; 2000US-228716P.

(CURA-) CURAGEN CORP.

Shinketsu RA, Leach MD;

WPI; 2002-106308/14.

P-PSDB; ABP02635.

Novel human polypeptides and polynucleotides useful for diagnosing, preventing and treating cardiovascular disease, neurodegenerative, hyperproliferative disorders and autoimmune disorders

Disclosure; SEQ ID 5251; 1037pp; English.

The present invention describes substantially purified human proteins (referred to as open reading frame, ORFX, where X is 1-1149) (see Table 1 in the specification). ABN15762 to ABN27252 encode the human ORFX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for treating or preventing a pathology associated with an ORFX-associated syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic

CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORF proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 379 BP; 80 A; 141 C; 94 G; 62 T; 2 other;

Query Match 57.6%; Score 19; DB 24; Length 379;
 Best Local Similarity 81.5%; Pred. No. 1e+02;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 OY 2 CCACATGAGTGGCAGCGCTCTGTGCA 28
 ||||| ||||| ||||| ||||| |||||
 DB 78 CCACATGCGTGGTGGTGGGCTCGGTGA 52

RESULT 8

AAS83486/C
 ID AAS83486 standard; cDNA; 841 BP.

AC AAS83486;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #19290.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG19299.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

PS Claim 1; SEQ ID NO 19290; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 841 BP; 191 A; 273 C; 224 G; 153 T; 0 other;

Query Match 57.0%; Score 18.8; DB 23; Length 841;
 Best Local Similarity 76.7%; Pred. No. 1.4e+02;
 Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 OY 3 CACATGAGTGGCAGCGCTGTGTATACC 32
 ||||| ||||| ||||| ||||| |||||
 DB 210 CAGGGGAGTGGCTTCTCTCTCTATACC 181

RESULT 9

ABN73219/C
 ID ABN73219 standard; cDNA; 472 BP.

AC ABN73219;

DT 03-JUL-2002 (first entry)

DE Bovine embryonic germ (EG) cell cDNA EST 000128a CONTIG 79.

KM Bovine; Bos taurus; EST; expressed sequence tag; totipotence;
 KM development; gene; ss.

OS Bos taurus.

PN WO200194550-A2.

PD 13-DEC-2001.

PF 07-JUN-2001; 2001WO-US18576.

PR 07-JUN-2000; 2000US-209874P.

PR 06-JUN-2001; 2001US-0876143.

PA (INF1-) INFGEN INC.

PI Elertsen KJ, Pfister-genskow M, Childs L;

DR WPI; 2002-351289/38.

PT An expressed sequence tag (EST), the expression of which, or its
 PT complementary sequence, in a cell identifies the cell as a
 PT developmentally competent or incompetent cell

PS Example 16; Page 129-130; 584pp; English.

CC The present invention describes an expressed sequence tag (EST), where
 CC the EST is an isolated, enriched, or purified nucleic acid sequence
 CC representing all or part of a gene, the expression of which, or its
 CC complementary sequence, in a cell identifies the cell as a
 CC developmentally competent or incompetent cell. Molecules which induce
 CC developmental competence in a cell line are useful for inducing
 CC totipotence in one or more cells. Molecules which induce developmental
 CC incompetence in a cell line are useful for preventing a full term
 CC pregnancy in an animal and inhibiting totipotence. The molecules are
 CC also useful for treating a disease in an animal by inducing development
 CC of one or more cells of the animal into a specific cell type. The
 CC present sequence represents a bovine EST which is given in the
 CC exemplification of the present invention.

Sequence 472 BP; 110 A; 132 C; 169 G; 57 T; 4 other;

Query Match 56.4%; Score 18.6; DB 24; Length 472;

Best Local Similarity 72.7%; Pred. No. 1.6e+02;

Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 402 GCCCATGCTGTCAGGCGCTGTGTACTG 370

1 GCCACATGATGTCAGGCGCTGTGTACTG 33

RESULT 10
AAC3442/C
AAC3442 standard; DNA; 973 BP.

AAC3442;

17-OCT-2000 (first entry)

Arabidopsis thaliana DNA fragment SEQ ID NO: 3049.

Hybridisation assay; genetic mapping; gene expression control;

protein identification; signal transduction pathway;

metabolic pathway; promoter; termination sequence; ss.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123348.

23-MAR-1999; 99US-0125788.

23-MAR-1999; 99US-0126264.

01-APR-1999; 99US-0126785.

06-APR-1999; 99US-0127462.

08-APR-1999; 99US-0128234.

16-APR-1999; 99US-0128714.

19-APR-1999; 99US-0128845.

21-APR-1999; 99US-0130077.

23-APR-1999; 99US-0130449.

23-APR-1999; 99US-0130510.

28-APR-1999; 99US-0130891.

30-APR-1999; 99US-0131448.

04-MAY-1999; 99US-0132048.

05-MAY-1999; 99US-0132407.

06-MAY-1999; 99US-0132484.

06-MAY-1999; 99US-0132486.

07-MAY-1999; 99US-0132487.

11-MAY-1999; 99US-0132863.

14-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134218.

14-MAY-1999; 99US-0134219.

14-MAY-1999; 99US-0134221.

18-MAY-1999; 99US-0134370.

19-MAY-1999; 99US-0134768.

20-MAY-1999; 99US-0134941.

21-MAY-1999; 99US-0135124.

24-MAY-1999; 99US-0135353.

25-MAY-1999; 99US-0135629.

27-MAY-1999; 99US-0136021.

28-MAY-1999; 99US-0136782.

01-JUN-1999; 99US-0137222.

03-JUN-1999; 99US-0137528.

04-JUN-1999; 99US-0137502.

07-JUN-1999; 99US-0137724.

08-JUN-1999; 99US-0138094.

10-JUN-1999; 99US-0138540.

10-JUN-1999; 99US-0138847.

14-JUN-1999; 99US-0139119.

16-JUN-1999; 99US-0139452.

16-JUN-1999; 99US-0139453.

17-JUN-1999; 99US-0139454.

18-JUN-1999; 99US-0139455.

18-JUN-1999; 99US-0139456.

18-JUN-1999; 99US-0139457.

18-JUN-1999; 99US-0139458.

18-JUN-1999; 99US-0139459.

18-JUN-1999; 99US-0139460.

18-JUN-1999; 99US-0139461.

18-JUN-1999; 99US-0139462.

18-JUN-1999; 99US-0139463.

18-JUN-1999; 99US-0139750.

21-JUN-1999; 99US-0139763.

22-JUN-1999; 99US-0139817.

23-JUN-1999; 99US-0139859.

24-JUN-1999; 99US-0140353.

28-JUN-1999; 99US-0140354.

29-JUN-1999; 99US-0140695.

30-JUN-1999; 99US-0140823.

01-JUL-1999; 99US-0140991.

01-JUL-1999; 99US-0141287.

02-JUL-1999; 99US-0141842.

06-JUL-1999; 99US-0142154.

06-JUL-1999; 99US-0142055.

08-JUL-1999; 99US-0142390.

09-JUL-1999; 99US-0142803.

12-JUL-1999; 99US-0142920.

13-JUL-1999; 99US-0142977.

14-JUL-1999; 99US-0143542.

15-JUL-1999; 99US-0143624.

16-JUL-1999; 99US-0144005.

16-JUL-1999; 99US-0144085.

19-JUL-1999; 99US-0144086.

19-JUL-1999; 99US-0144325.

19-JUL-1999; 99US-0144331.

19-JUL-1999; 99US-0144332.

19-JUL-1999; 99US-0144333.

19-JUL-1999; 99US-0144334.

20-JUL-1999; 99US-0144335.

20-JUL-1999; 99US-0144352.

21-JUL-1999; 99US-0144632.

21-JUL-1999; 99US-0144884.

21-JUL-1999; 99US-0145086.

22-JUL-1999; 99US-0145088.

22-JUL-1999; 99US-0145089.

22-JUL-1999; 99US-0145087.

22-JUL-1999; 99US-0145089.

23-JUL-1999; 99US-0145128.

23-JUL-1999; 99US-0145218.

26-JUL-1999; 99US-0145224.

27-JUL-1999; 99US-0145276.

27-JUL-1999; 99US-0145913.

27-JUL-1999; 99US-0145918.

28-JUL-1999; 99US-0145919.

02-AUG-1999; 99US-0145951.

02-AUG-1999; 99US-0146386.

02-AUG-1999; 99US-0146388.

03-AUG-1999; 99US-0146389.

04-AUG-1999; 99US-0147038.

04-AUG-1999; 99US-0147204.

05-AUG-1999; 99US-0147302.

05-AUG-1999; 99US-0147192.

06-AUG-1999; 99US-0147260.

06-AUG-1999; 99US-0147303.

09-AUG-1999; 99US-0147416.

09-AUG-1999; 99US-0147493.

```

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155119.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 56.4%; Score 18.6; DB 21; Length 973;
Best Local Similarity 72.7%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 GCCACATGAGTGGCAAGCGCTGTGTGATACCG 33

```

```

Db 268 GCACAGAGAGCAAGAGAGCTGTGTATCAG 236
|||||
RESULT 11
AAS83436/c
ID AAS83436 standard; cDNA; 1122 BP.
XX
AC AAS83436;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #19240.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
MO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HXSE-) HXSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
PI WPI; 2001-639362/73.
XX
DR P-PSDB; ABG19249.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1; SEQ ID NO 19240; 103pp; English.
XX
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp://wipo.int/pub/published_pcr_sequences.
XX
SQ Sequence 1122 BP; 216 A; 400 C; 309 G; 196 T; 1 other;

```

```

Query Match 56.4%; Score 18.6; DB 23; Length 1122;
Best Local Similarity 72.7%; Pred. No. 1.8e+02;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 GCCACATGAGTGGCAAGCGCTGTGTATACCG 33
|||||
Db 1074 GCACAGAGAGAGCAAGAGAGCTGTGTATACCG 1042

```

RESULT 12

ABN85733/c
ID: ABN85733 standard; DNA: 249487 BP.

ABN85733;

30-SEP-2002 (first entry)

Mouse genomic region containing the Itprcs gene SEQ ID NO 3.

Mouse; Itprcs; taste; cell signalling; TC-ICS; food; pharmaceutical; taste cell-specific ion channel subunit; gene; ds.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Disclosure; Page 64-207; 306pp; English.

The invention relates to identifying (MI) a compound that modulates taste signalling in taste cells, by contacting the compound with a eukaryotic host cell or cell membrane which expresses a taste cell-specific ion channel subunit (TC-ICS), and determining a functional effect of the compound upon a transmembrane ion flux of a predetermined ion.

(MI) is useful for identifying a compound that modulates taste signalling in taste cells, for identifying a compound that binds to a taste cell specific ion channel subunit and for modulating taste signalling in taste cells of a mammal, in particular a human. Modulators identified by (MI) are used by the food and pharmaceutical industries to customize taste, e.g. as additives to food or medicine so that the food or medicine tastes different to the subject who ingests it. Bitter medicines can be made to taste less bitter and sweet substance can be enhanced. The modulators are useful for pharmacological and genetic modulation of taste signalling pathways. The taste modulators can be directly administered to mammalian subjects for modulation of taste in vivo. The present sequence is that of the mouse genomic region containing the Itprcs gene of the invention.

Sequence 249487 BP; 64808 A; 61913 C; 61498 G; 61236 T; 32 other;

Query Match 56.4%; Score 18.6; DB 24; Length 249487; Best Local Similarity 84.0%; Pred. No. 3.8e+02; Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

4 ACATGAGTGGCAAGCGCTCTGCGA 28
|||||
25090 ACAGAGTGGCAAGCGTATGGCGA 25066

RESULT 13
AAA96719/c
ID AAA96719 standard; DNA: 1131 BP.

AC AAA96719;

19-FEB-2001 (first entry)

Polynucleotide isolated from lymph node stromal cells of f3n -/- mice.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Polypeptide expressed in mammalian f3n -/- lymph node stromal cells, useful for modulating growth of blood cells, for treating inflammatory and tumour necrosis factor-mediated disorders, cancer and viral disorders

Claim 4; Page 35; 75pp; English.

The present sequence represents a polynucleotide sequence which is isolated from lymph node stromal cells of f3n -/- mice. The polynucleotides and their polypeptides are useful for treating an inflammatory disorder, disorder of immune system and cancer selected from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a viral disorder, in particular HIV infection and for modulating the growth of blood vessels. The polypeptides are useful for treating a tumour necrosis factor (TNF) mediated disorder, such as those selected from arthritis, inflammatory bowel disease and cardiac failure and a fibroblast growth factor-mediated disorder. It is also useful in assays to determine biological activity, to raise antibodies, to isolate corresponding ligands or receptors, to quantify levels of protein or and in compositions for the treatment of skin, connective tissue and immune system diseases. The polynucleotide is useful as marker for genetic disorder.

Sequence 1131 BP; 208 A; 343 C; 342 G; 238 T; 0 other;

Query Match 55.8%; Score 18.4; DB 21; Length 1131; Best Local Similarity 78.6%; Pred. No. 2.2e+02; Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

1 GCCACATGAGTGGCAAGCGCTCTGCGA 28
|||||
435 GCCAGATGAGTGGCAAGCGCTCTGCGA 408

RESULT 14

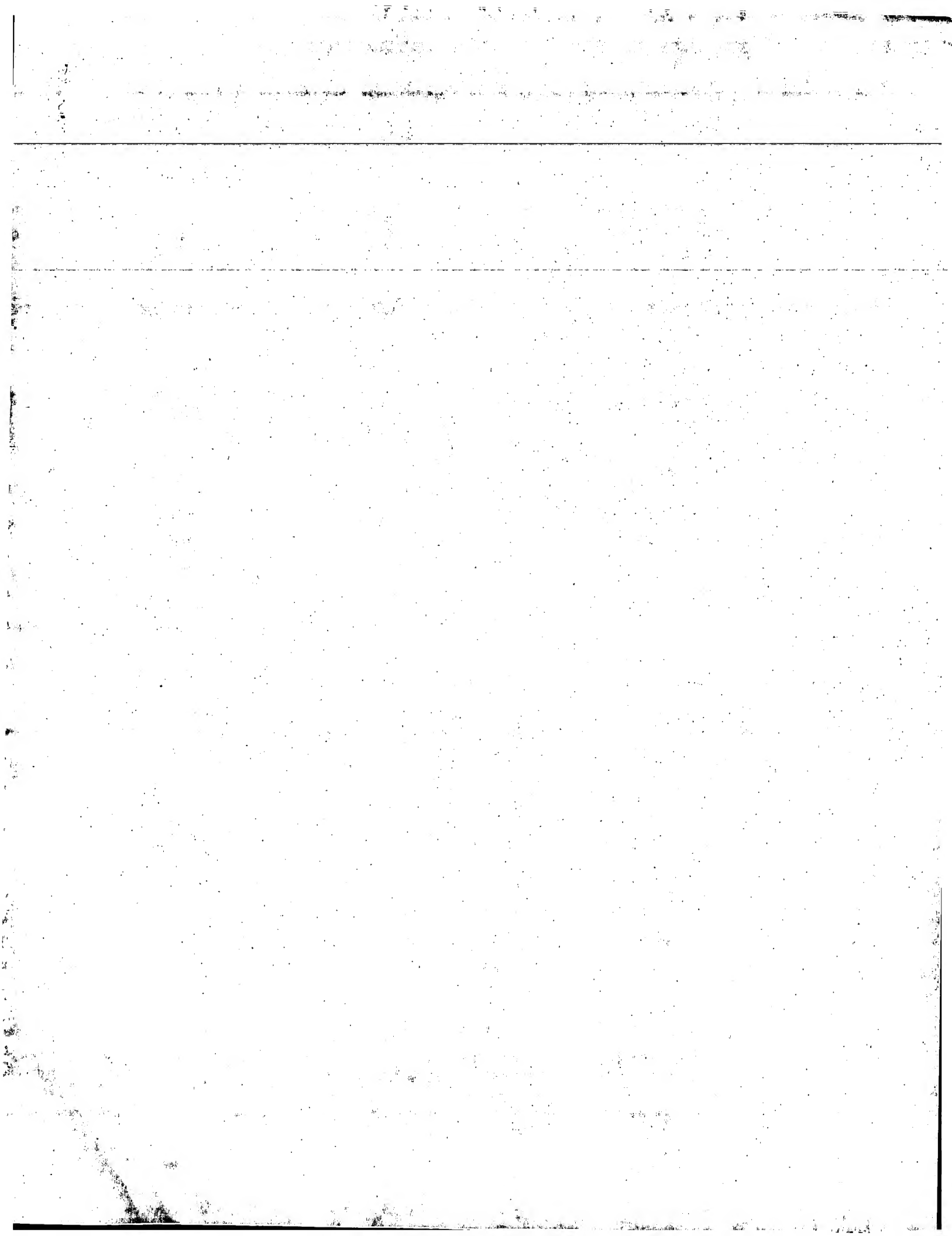
AAF93747

ID AAF93747 standard; cDNA; 3425 BP.
 XX AAF93747;
 AC
 XX
 DT 23-MAY-2001 (first entry)
 XX
 XX
 DE Human cDNA encoding a membrane or secretory protein clone PSEC0007.
 XX
 XX
 KM Human; secretory protein; membrane protein; vaccine; gene therapy;
 KM rheumatoid arthritis; diabetes; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1067182-A2.
 XX
 PD 10-JAN-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114090.
 XX
 PR 08-JUL-1999; 99JP-0194179.
 PR 11-JAN-2000; 2000JP-0118775.
 PR 02-MAY-2000; 2000JP-0183766.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 XX
 DR WPI: 2001-093989/11.
 DR P-PSDB: AAB88320.
 XX
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
 PT gene therapy or as candidate target molecules in drug development -
 XX
 PS Claim 1; SEQ ID 7; 609bp + CD ROM; English.
 XX
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by
 CC AAB88317 - AAB88419. Included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be used
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies (agonists and antagonists) of expression and
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
 CC (ELISA). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.
 CC
 SQ Sequence 3425 BP; 800 A; 840 C; 917 G; 868 T; 0 other;
 XX
 XX
 Query Match 55.8%; Score 18.4; DB 22; Length 3425;
 Best Local Similarity 78.6%; Pred. No. 2.5e+02;
 Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 OY 1 GCCACATGATGGCAAGGCGTCTGCTGA 28
 DB 1891 GCCACCTGTGTGGCAAGGCGTCACTGA 1918

RESULT 15
 AA248806

ID AA248806 standard; cDNA; 518 BP.
 XX AA248806;
 AC
 XX
 DT 21-MAR-2000 (first entry)
 XX
 XX
 DE Rice inositol 1,3,4-triphosphate 5/6-kinase coding sequence.
 XX
 XX
 KM Inositol 1,3,4-triphosphate 5/6-kinase; phytic acid; genetic mapping;
 KM myo-inositol 1,2,3,4,5,6-hexaphosphate; biosynthetic enzyme; phytate;
 KM animal feed; ss.
 XX
 OS Oryza sativa.
 XX
 PN WO9955879-A1.
 XX
 PD 04-NOV-1999.
 XX
 PF 22-APR-1999; 99WO-US08790.
 XX
 PR 24-APR-1998; 98US-0082960.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Cahoon RE, Carlson TJ, Hiltz WD, Pearlstein RM;
 XX
 DR WPI: 2000-072179/06.
 DR P-PSDB: AAV59422.
 XX
 PT Novel phytic acid biosynthetic enzymes used to alter the level of the
 PT enzyme in transgenic plants -
 XX
 PS Claim 2; Page 45; 63bp; English.
 XX
 CC This sequence encodes an inositol 1,3,4-triphosphate 5/6-kinase (ITK)
 CC protein of the invention. ITK is a phytic acid (myo-inositol
 CC 1,2,3,4,5,6-hexaphosphate) biosynthetic enzyme. The ITK enzymes of the
 CC invention may be prepared recombinantly and used to raise antibodies,
 CC which are used for detecting the enzymes in situ in cells or in vitro
 CC in cell extracts. The polynucleotides of the invention may be used to
 CC create transgenic plants in which the ITK levels are present at higher
 CC or lower levels than normal, or in cell types or developmental processes
 CC where they are not normally found. This would alter the level of
 CC 1,3,4-triphosphate 5/6 kinase found in those cells. In addition, it may
 CC be desirable to eliminate expression of ITK genes for certain
 CC applications. The polynucleotides also serve as a source of probes and
 CC primers, which are useful for genetic mapping, as markers for traits
 CC linked to those genes, and to isolate homologous sequences from other
 CC species, as well as for physical mapping. Decreased amounts of phytate
 CC and increased amounts of available phosphate in animal feeds such as corn
 CC would lead to improved feed efficiency. The proteins of the present
 CC invention lead to a better understanding of the phytic acid biosynthesis
 CC pathway, allowing it to be exploited for commercial uses, e.g. in animal
 CC feeds.
 CC
 SQ Sequence 518 BP; 120 A; 114 C; 117 G; 159 T; 8 other;
 XX
 XX
 Query Match 55.2%; Score 18.2; DB 21; Length 518;
 Best Local Similarity 87.0%; Pred. No. 2.4e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 5 CATGATGTCGCAAGCGCTGCTG 27
 DB 198 CATGTGTGTCGCAAGCGCTGCTG 220

Search completed: July 10, 2003, 19:52:39
 Job time : 72.6337 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 ; Search time 14.7573 Seconds
(without alignments)
685.785 Million cell updates/sec

Title: US-10-087-631b-5

Perfect score: 33

Sequence: 1 gccacatgagtggaagcgctcgtgatccg 33

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description
	1	18.4	55.8	320	4	US-08-651-155B-214
	2	18.4	55.8	1131	4	US-09-383-586-7
	3	18.2	55.2	1740	2	US-08-471-025-37
	4	18.2	55.2	7785	2	US-08-276-967-1
	5	17.8	53.9	617	4	US-09-385-982-396
	6	17.8	53.9	1279	2	US-08-872-719-1
	7	17.8	53.9	1279	3	US-08-957-302A-11
	8	17.8	53.9	1279	4	US-09-336-890-1
	9	17.8	53.9	1279	4	US-09-542-403-11
	10	17.8	53.9	1279	4	US-09-668-499-1
	11	17.8	53.9	1395	3	US-08-957-302A-3
	12	17.8	53.9	1395	4	US-08-542-403-3
	13	17.8	53.9	1575	3	US-08-957-302A-1
	14	17.8	53.9	1575	4	US-09-542-403-1
	15	17.6	53.3	949	4	US-09-221-017B-615
	16	17.6	53.3	2532	1	US-07-671-376C-4
	17	17.6	53.3	19227	3	US-09-090-793-13
	18	17.6	53.3	40138	3	US-09-090-793-12
	19	17.4	52.7	456	1	US-08-686-878A-19
	20	17.4	52.7	456	1	US-09-175-928-19
	21	17.4	52.7	2805	4	US-08-959-004-6
	22	17.2	52.1	4403765	4	US-09-103-840A-2
	23	17.2	52.1	4411529	4	US-09-103-840A-1
	24	17.2	51.5	51	1	US-08-396-218-3
	25	17.2	51.5	51	1	US-08-396-218-3
	26	17.2	51.5	59	1	US-08-396-218-4
	27	17.2	51.5	59	1	US-08-760-116-4

c 28	17	51.5	318	1	US-08-259-372A-15	Sequence 15, Appl
c 30	17	51.5	318	1	US-08-468-671-15	Sequence 15, Appl
c 32	17	51.5	594	4	US-09-385-982-397	Sequence 397, App
c 31	17	51.5	620	1	US-08-328-055-2	Sequence 2, Appl
c 33	17	51.5	642	2	US-08-480-753-5	Sequence 10, Appl
c 34	17	51.5	642	3	US-09-041-889-10	Sequence 10, Appl
c 35	17	51.5	785	1	US-08-837-058-10	Sequence 10, Appl
c 36	17	51.5	1446	1	US-08-596-024-5	Sequence 5, Appl
c 37	17	51.5	1446	4	US-09-020-818-5	Sequence 5, Appl
c 38	17	51.5	1446	4	US-08-907-740-5	Sequence 5, Appl
c 39	17	51.5	2081	2	US-09-096-982-7	Sequence 7, Appl
c 40	17	51.5	2081	2	US-08-653-650A-7	Sequence 31, Appl
c 41	17	51.5	2910	4	US-09-206-942-31	Sequence 29, Appl
c 42	17	51.5	2928	4	US-09-206-942-29	Sequence 29, Appl
c 43	17	51.5	3196	2	US-09-096-982-4	Sequence 4, Appl
c 44	17	51.5	3196	2	US-08-653-650A-4	Sequence 4, Appl
c 45	17	51.5	13875	2	US-08-734-344-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-651-155B-214
Sequence 214, Application US/08651155B
Patent No. 6365401
GENERAL INFORMATION:
APPLICANT: Mahan Dr., Michael J.
APPLICANT: Conner Mr., Christopher P.
APPLICANT: Hiethoff Mr., Douglas M.
TITLE OF INVENTION: METHOD AND PROBS FOR THE IDENTIFICATION
TITLE OF INVENTION: OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST
NUMBER OF SEQUENCES: 255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chrisman, Bynum & Johnson, P.C.
STREET: 1900 Fifteenth Street
CITY: Boulder
STATE: CO
COUNTRY: USA
ZIP: 80302
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/651,155B
FILING DATE: 17-MAY-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Petersen Mr., Steven C.
REGISTRATION NUMBER: 36,238
REFERENCE/DOCKET NUMBER: 17060.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/546-1300
TELEFAX: 303/449-5426
TELEX: ABA1475
INFORMATION FOR SEQ ID NO: 214:
SEQUENCE CHARACTERISTICS:
LENGTH: 320 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: YES
US-08-651-155B-214
Query Match 55.8%; Score 18.4; DB 4; Length 320;
Best Local Similarity 78.6%; Pred. No. 19;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 6 ATGAGTGCAGAGCGCTGCGATACG.33
111 1111 111111 11 1111
Db 196 ATGGCTGGCATGAGCTGCGCAACCG 223

RESULT 2
US-09-383-586-7/c

Sequence 7, Application US/09383586
Patent No. 6242419

GENERAL INFORMATION:

APPLICANT: Strachan, Lorna

APPLICANT: Sleeman, Matthew

APPLICANT: Abernethy, Nevin

APPLICANT: Onrust, Rene

APPLICANT: Kuntze, Anand

APPLICANT: Murtison, Greg

TITLE OF INVENTION: Compounds isolated from stromal cells

FILE REFERENCE: 11000.1037c1

CURRENT APPLICATION NUMBER: US/09/383,586

NUMBER OF SEQ ID NOS: 38

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 7

LENGTH: 1131

TYPE: DNA

ORGANISM: Mouse

US-09-383-586-7

Query Match

Best Local Similarity 55.88; Score 18.4; DB 4; Length 1131;

Matches 32; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

1 GCCACATGAGTGCAGAGCGCTGCGTGA 28

435 GCCAGATGAGTGCAGAGCGCTGCGCA 408

Db

RESULT 3

US-08-471-025-37/c

Sequence 37, Application US/08471025

Patent No. 598561

GENERAL INFORMATION:

APPLICANT: Paolletti, Enzo

APPLICANT: Fischer, Laurent

TITLE OF INVENTION: RECOMBINANT POXVIRUS - CALICIVIRUS

TITLE OF INVENTION: (RABBIT HEMORRHAGIC DISEASE VIRUS (RHDV)) COMPOSITIONS AND

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESS: Curtis, Morris & Safford, P.C.

STREET: 530 Fifth Avenue, 25th Floor

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471,025

FILING DATE: 06-JUN-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Frommer, William S.

REGISTRATION NUMBER: 25,506

REFERENCE/DOCKET NUMBER: 454310-2650

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 840-3333

TELEFAX: (212) 840-0712

INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:

LENGTH: 1740 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-471-025-37

Query Match

Best Local Similarity 55.28; Score 18.2; DB 2; Length 1740;

Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 CCACATGAGTGCAGAGCGCTGCGTACG 32

Db 1020 CCACATGAGTGCAGAGCGCTGCGTACG 990

RESULT 4

US-08-276-967-1

Sequence 1, Application US/08276967

Patent No. 5851817

GENERAL INFORMATION:

APPLICANT: Hardy, Daniel M.

APPLICANT: Garbers, David L.

TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of

TITLE OF INVENTION: Sperm

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P. O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/276,967

FILING DATE: Submitted Herewith

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Kitchell, Barbara S.

REGISTRATION NUMBER: 33,928

REFERENCE/DOCKET NUMBER: UTSD:418\KIT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 713-787-1400

TELEFAX: 713-789-2679

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 7785 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-276-967-1

Query Match

Best Local Similarity 55.28; Score 18.2; DB 2; Length 7785;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 10 GTGGCAGAGCGCTGCGTATACC 32

Db 509 GTGGCAGAGCGCTGCGTATACC 531

RESULT 5

US-09-385-982-396

Sequence 396, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCDA-260XX
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSeq for Windows: Version 3.0
SEQ ID NO 396
LENGTH: 617
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(617)
OTHER INFORMATION: n - A,T,C or G
US-09-385-982-396

Query Match
Best Local Similarity 53.9%; Score 17.8; DB 4; Length 617;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 GCCACATGAGTGGCAAGCGCTGTGTGAT 29
DB 12 GCCACTGAGTGGCAAGCGCTGTGTGAT 40

RESULT 6
US-08-872-719-1
Sequence 1, Application US/08872719
Patent No. 5919643
GENERAL INFORMATION:
APPLICANT: Kelley, Mark R.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE USE OF APURINIC/APYRIMIDINIC
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77057-2196
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/872,719
FILING DATE: CONCURRENTLY HERewith
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: INDY:012P21
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1279 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-872-719-1

Query Match
Best Local Similarity 53.9%; Score 17.8; DB 2; Length 1279;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GCCACATGAGTGGCAAGCGCTGTGTGAT 29
DB 821 GCCACAAGAGCGCCCAAGCGCTGTGTGAT 849

RESULT 7

US-08-957-302A-11
Sequence 11, Application US/08957302A
Patent No. 6046036

GENERAL INFORMATION:
APPLICANT: Williams, David
TITLE OF INVENTION: DNA Sequences Encoding Fusions of DNA
TITLE OF INVENTION: Repair Proteins and Uses Thereof
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TX
COUNTRY: US

ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,302A
FILING DATE: Concurrently Herewith
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: INDY:005
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1279 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-957-302A-11

Query Match
Best Local Similarity 53.9%; Score 17.8; DB 3; Length 1279;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GCCACATGAGTGGCAAGCGCTGTGTGAT 29
DB 821 GCCACAAGAGCGCCCAAGCGCTGTGTGAT 849

RESULT 8

US-09-336-890-1
Sequence 1, Application US/09336890
Patent No. 6190661

GENERAL INFORMATION:
APPLICANT: Kelley, Mark R.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE USE OF
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1279 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-542-403-11

Query Match
Best Local Similarity 53.9%; Score 17.8; DB 4; Length 1279;
Pred. No. 45;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GCCACATGAGTGGCAGCGCTGTGTAT 29
||||| ||| ||||| || || ||
821 GCCACAGAGCGCCAGGCTTCGGGAA 849

RESULT 10
US-09-668-499-1
Sequence 1, Application US/09668499
Patent No. 6406917
GENERAL INFORMATION:
APPLICANT: Kelley, Mark R.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE USE OF APURINIC/APYRIMIDIN
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77057-2198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/668,499
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/872,719
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: INDY:012P21
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1279 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-668-499-1

Query Match
Best Local Similarity 53.9%; Score 17.8; DB 4; Length 1279;
Pred. No. 45;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GCCACATGAGTGGCAGCGCTGTGTAT 29
||||| ||| ||||| || || ||
821 GCCACAGAGCGCCAGGCTTCGGGAA 849

```

Sequence 3, Application US/08957302A
Patent No. 6046036
GENERAL INFORMATION:
APPLICANT: Kelley, Mark
TITLE OF INVENTION: DNA Sequences Encoding Fusions of DNA
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESSES:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TX
COUNTRY: US
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,302A
FILING DATE: Concurrently Herewith
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: INDY:005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1395 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-957-302A-3

Query Match 53.9%; Score 17.8; DB 3; Length 1395;
Best Local Similarity 75.9%; Pred. No. 46;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GCCACATGAGTGGCAAGCGCTGCTGTGAT 29
||||| ||| | ||||| ||| ||
Db 1137 GCCACAAGAGCGGCCAAGCGCTCGGGGAT 1165

RESULT 12
US-09-542-403-3
Sequence 3, Application US/09542403
Patent No. 6252048
GENERAL INFORMATION:
APPLICANT: Kelley, Mark
TITLE OF INVENTION: DNA Sequences Encoding Fusions of DNA
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESSES:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TX
COUNTRY: US
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/542,403
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/957,302
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: INDY:005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1395 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-542-403-3

Query Match 53.9%; Score 17.8; DB 4; Length 1395;
Best Local Similarity 75.9%; Pred. No. 46;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GCCACATGAGTGGCAAGCGCTGCTGTGAT 29
||||| ||| | ||||| ||| ||
Db 1137 GCCACAAGAGCGGCCAAGCGCTCGGGGAT 1165

RESULT 13
US-08-957-302A-1
Sequence 1, Application US/08957302A
Patent No. 6046036
GENERAL INFORMATION:
APPLICANT: Kelley, Mark
TITLE OF INVENTION: DNA Sequences Encoding Fusions of DNA
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESSES:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TX
COUNTRY: US
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,302A
FILING DATE: Concurrently Herewith
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: INDY:005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1575 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-957-302A-1

Query Match 53.9%; Score 17.8; DB 3; Length 1575;
Best Local Similarity 75.9%; Pred. No. 47;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 GCCACATGAGTGGCAGGCGTGTGAT 29
DB 1317 GCCACAGAGCGCCAGGCTTCGGGAAAT 1345

RESULT 14

US-09-542-403-1

Sequence 1, Application US/09542403

Patent No. 6252048

GENERAL INFORMATION:

APPLICANT: Kelley, Mark

APPLICANT: Williams, David

TITLE OF INVENTION: DNA Sequences Encoding Fusions of DNA

TITLE OF INVENTION: Repair Proteins and Uses Thereof

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESSES:

ADDRESSEE: ARNOLD, WHITE & DUNKER

STREET: P.O. BOX 4433

CITY: HOUSTON

STATE: TX

COUNTRY: US

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/542,403

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/957,302

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Highlander, Steven L.

REGISTRATION NUMBER: 37,642

REFERENCE/DOCKET NUMBER: INDY:005

TELEPHONE: (512) 418-3000

TELEFAX: (713) 789-2679

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1575 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-542-403-1

Query Match

Best Local Similarity

Matches

OY 1 GCCACATGAGTGGCAGGCGTGTGAT 29

DB 1317 GCCACAGAGCGCCAGGCTTCGGGAAAT 1345

US-09-221-017B-615

Sequence 615, Application US/09221017B

Patent No. 6444799

GENERAL INFORMATION:

APPLICANT: Ross, Bruce C.

TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 PAGE MILL ROAD

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FASTSEQ for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/221,017B

FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: MONROY, Gladys H

REGISTRATION NUMBER: 32,430

REFERENCE/DOCKET NUMBER: 27340-20021.00

TELEPHONE: 650-813-5600

TELEFAX: 650-494-0792

INFORMATION FOR SEQ ID NO: 615:

SEQUENCE CHARACTERISTICS:

LENGTH: 949 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: UNKNOWN

ORIGINAL SOURCE:

ORGANISM: PORPHYROMONAS GINGIVALIS

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1...949

US-09-221-017B-615

Query Match

Best Local Similarity

Matches

OY 1 GCCACATGAGTGGCAGGCGTGTGATAC 32

DB 829 GCATATGATTTGCCAGAGATTTCGGGATAC 860

US-09-221-017B-615

Sequence 615, Application US/09221017B

Patent No. 6444799

GENERAL INFORMATION:

APPLICANT: Ross, Bruce C.

TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 PAGE MILL ROAD

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

Search completed: July 10, 2003, 20:27:34

Job time: 34.7573 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:53:16 ; Search time 90.3236 Seconds
(without alignments)
576.891 Million cell updates/sec

Title: US-10-087-631B-5
Perfect score: 33
Sequence: 1 gccacatgagtgcaagcgctgtgatccg 33

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications, NA:*

- 1: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
- 2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 6: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/prodata/1/pubpna/US60_PUB_PUB.seq:*
- 14: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	33	100.0	33	9	US-10-087-631B-5
2	22	66.7	241	9	US-10-087-631B-12
3	22	66.7	242	9	US-10-087-631B-15
4	22	66.7	375	9	US-10-087-631B-14
5	22	66.7	943	9	US-10-087-631B-11
6	20	60.6	744	10	US-09-995-598-85
7	20	60.6	744	10	US-09-995-598-87
8	20	60.6	744	10	US-09-995-598-89
9	19.8	60.0	771	9	US-10-013-173-38
10	19.8	60.0	771	9	US-10-150-762-38
11	19.4	58.8	499	9	US-09-918-995-25110
12	19.2	58.2	405	10	US-09-960-352-13265
13	19	57.6	161280	9	US-10-144-649A-746
14	18.6	56.4	1197	9	US-10-156-761-409
15	18.6	56.4	100000	9	US-10-156-761-15103
16	18.6	56.4	249487	9	US-10-026-188-3
17	18.6	56.4	9025608	9	US-10-156-761-1
18	18.4	55.8	1131	10	US-09-823-038A-7
19	18.2	55.2	189	10	US-09-783-590-9470

Query Match	Score	DB	Length	Matches	Conservative	Mismatches	Indels	Gaps
1	100.0%	33	33	33	0	0	0	0
2	100.0%	33	33	33	0	0	0	0
3	100.0%	33	33	33	0	0	0	0
4	100.0%	33	33	33	0	0	0	0
5	100.0%	33	33	33	0	0	0	0
6	100.0%	33	33	33	0	0	0	0
7	100.0%	33	33	33	0	0	0	0
8	100.0%	33	33	33	0	0	0	0
9	100.0%	33	33	33	0	0	0	0
10	100.0%	33	33	33	0	0	0	0
11	100.0%	33	33	33	0	0	0	0
12	100.0%	33	33	33	0	0	0	0
13	100.0%	33	33	33	0	0	0	0
14	100.0%	33	33	33	0	0	0	0
15	100.0%	33	33	33	0	0	0	0
16	100.0%	33	33	33	0	0	0	0
17	100.0%	33	33	33	0	0	0	0
18	100.0%	33	33	33	0	0	0	0
19	100.0%	33	33	33	0	0	0	0

ALIGNMENTS

RESULT 1

US-10-087-631B-5
Sequence 5, Appli Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087, 631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 33
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: ST650pc probe sequence
US-10-087-631B-5

Query Match 100.0%; Score 33; DB 9; Length 33;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2

US-10-087-631B-12/C
Sequence 12, Appli Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087, 631B
CURRENT FILING DATE: 2002-03-01

NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 12
LENGTH: 241
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Amplicon derived from QS (pc
US-10-087-631B-12

Query Match 66.7%; Score 22; DB 9; Length 241;
Best Local Similarity 97.1%; Pred. No. 1.3;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GCCACATGAGTGG-CAAGCGCTGTGTGATACCG 33
103 GCCACATGAGTGGCCAAAGCGCTGTGTGATACCG 70

RESULT 3
US-10-087-631B-15/c
Sequence 15, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 15
LENGTH: 242
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of artificial sequence: Amplicon derived from ICSJ62
US-10-087-631B-15

Query Match 66.7%; Score 22; DB 9; Length 242;
Best Local Similarity 97.1%; Pred. No. 1.3;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GCCACATGAGTGG-CAAGCGCTGTGTGATACCG 33
103 GCCACATGAGTGGCCAAAGCGCTGTGTGATACCG 70

RESULT 4
US-10-087-631B-14/c
Sequence 14, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 375
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: ICSJ620 HCV (HCV specific
OTHER INFORMATION: amplification control having a binding site for S7280 and S7778
US-10-087-631B-14

Query Match 66.7%; Score 22; DB 9; Length 375;
Best Local Similarity 97.1%; Pred. No. 1.3;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GCCACATGAGTGG-CAAGCGCTGTGTGATACCG 33
176 GCCACATGAGTGGCCAAAGCGCTGTGTGATACCG 143

RESULT 5
US-10-087-631B-11/c
Sequence 11, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 11
LENGTH: 943
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: QS (pc) HCV being parallel
US-10-087-631B-11

Query Match 66.7%; Score 22; DB 9; Length 943;
Best Local Similarity 97.1%; Pred. No. 1.4;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GCCACATGAGTGG-CAAGCGCTGTGTGATACCG 33
159 GCCACATGAGTGGCCAAAGCGCTGTGTGATACCG 126

RESULT 6
US-09-995-598-85/c
Sequence 85, Application US/09995598
Patent No. US20020142293A1
GENERAL INFORMATION:
APPLICANT: CARINO, RADU
APPLICANT: GUILLOT, SOPHIE
TITLE OF INVENTION: ENTEROVIURAL POLYNUCLEOTIDES, METHODS OF DETECTING ENTEROVIURSE
FILE REFERENCE: 215059US-660-660-0
CURRENT APPLICATION NUMBER: US/09/995,598
CURRENT FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: US 60/253741
PRIOR FILING DATE: 2000-11-29
NUMBER OF SEQ ID NOS: 96
SOFTWARE: PatentIn version 3.1
SEQ ID NO 85
LENGTH: 744
TYPE: DNA
ORGANISM: Echovirus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(744)
OTHER INFORMATION:
US-09-995-598-85

Query Match 60.6%; Score 20; DB 10; Length 744;
Best Local Similarity 82.1%; Pred. No. 11;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 CATGAGTGGCAAGCGCTGTGTGATACCG 32

Db 54 CAAGATCAGCTGCAGGCTTCTGTTATACC 84

RESULT 11

US-09-918-995-25170
; Sequence 25170, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918, 995
; PRIOR FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 25170
; LENGTH: 499
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)...(499)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-25170

Query Match 58.8%; Score 19.4; DB 9; Length 499;
Best Local Similarity 79.3%; Pred. No. 21;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 455 ACATAGTGTGCGAGCGCTGTGATACC 32
ACATAGTGTGCGAGCGCTGTGATACC 483

RESULT 12

US-09-960-352-13265
; Sequence 13265, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Ningbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511, 006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; PRIOR FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO: 13265
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 57-LIB188-003-Q1-E1-G10
US-09-960-352-13265

Query Match 58.2%; Score 19.2; DB 10; Length 405;
Best Local Similarity 75.0%; Pred. No. 25;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 113 GCCATATGAGTGTGCGAGCGCTGTGATACC 32
GCCATATGAGTGTGCGAGCGCTGTGATACC 144

RESULT 13

US-10-144-649A-746
; Sequence 746, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:

APPLICANT: Lodes, Michael J.
APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Algate, Paul A.
APPLICANT: Monell, Patricia D.
TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121, 475C11
CURRENT APPLICATION NUMBER: US/10/144, 649A
PRIOR FILING DATE: 2002-08-21
NUMBER OF SEQ ID NOS: 749
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 746
LENGTH: 161280
TYPE: DNA
ORGANISM: Homo sapiens
US-10-144-649A-746

Query Match 57.6%; Score 19; DB 9; Length 161280;
Best Local Similarity 81.5%; Pred. No. 48;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 2 CCACATGAGTGTGCGAGCGCTGTGATACC 28
CCACATGAGTGTGCGAGCGCTGTGATACC 39987

RESULT 14

US-10-156-761-409/C
; Sequence 409, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO: 409
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1197)
US-10-156-761-409

Query Match 56.4%; Score 18.6; DB 9; Length 1197;
Best Local Similarity 72.7%; Pred. No. 51;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 1019 GCCACATGAGTGTGCGAGCGCTGTGATACC 33
GCCACATGAGTGTGCGAGCGCTGTGATACC 987

RESULT 15

US-10-156-761-15103/C
; Sequence 15103, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO

: APPLICANT: ISHIKAWA, JUN
 : APPLICANT: HORIKAWA, HIROSHI
 : APPLICANT: SHIBA, TADAYOSHI
 : APPLICANT: SAKAKI, YOSHIYUKI
 : APPLICANT: HATORI, MASAHIRA
 : TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 : FILE REFERENCE: 249-262
 : CURRENT APPLICATION NUMBER: US/10/156,761
 : CURRENT FILING DATE: 2002-05-29
 : PRIOR APPLICATION NUMBER: JP 2001-204089
 : PRIOR FILING DATE: 2001-05-30
 : PRIOR APPLICATION NUMBER: JP 2001-272697
 : PRIOR FILING DATE: 2001-08-02
 : NUMBER OF SEQ ID NOS: 15109
 : SEQ ID NO 15103
 : LENGTH: 100000
 : TYPE: DNA
 : ORGANISM: Streptomyces avermitilis
 US-10-156-761-15103

Query Match 56.4%; Score 18.6; DB 9; Length 100000;
 Best Local Similarity 72.7%; Pred. No. 71;
 Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GCCACATGAGTGGCAGCGCTGTGTGATACCG 33
 ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 81448 GCCAGATGCGGGCGGCGGCGCTGCGATGTCG 81416

Search completed: July 11, 2003, 15:01:56
 Job time: 100.324 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:13:45 ; Search time 593.333 Seconds
(without alignments)
900.761 Million cell updates/sec

Title: US-10-087-631B-5
Perfect score: 33
Sequence: 1 gccacatgagtgcaagcgctctgtgatccg 33

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estopl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rnd:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22.6	68.5	615	10	AV736417 AV736417
2	22.4	67.9	587	10	AM180058 MGN0121f
3	22	66.7	211	9	AV003500 AV003500
4	22	66.7	593	17	AQ989608 RfC00171
5	21.6	65.5	118	10	AM804244 PM3-DK008
6	21.4	64.8	149	17	AZ922863 STC05F03

c	7	21.2	64.2	777	12	BE676553
c	8	21	63.6	230	10	BE078191
c	9	21	63.6	609	10	AV744190
c	10	21	63.6	778	12	BE784735
c	11	20.8	63.0	538	13	B1296816
c	12	20.6	62.4	601	10	AV737560
c	13	20.6	62.4	602	10	AV744188
c	14	20.6	62.4	613	10	AV737831
c	15	20.6	62.4	777	9	A1525664
c	16	20.6	62.4	960	12	B6112389
c	17	20.4	61.8	501	10	BE375363
c	18	20.4	61.8	589	14	BQ321112
c	19	20.2	61.2	337	9	AA074362
c	20	20.2	61.2	496	10	AV643583
c	21	20.2	61.2	618	9	A1854588
c	22	20.2	61.2	699	9	AA089186
c	23	20.2	61.2	1062	12	BE799927
c	24	20.2	61.2	484	12	BG886095
c	25	20	60.6	489	17	BH331185
c	26	20	60.6	561	9	AL794328
c	27	20	60.6	605	9	AL794563
c	28	20	60.6	628	9	AL782209
c	29	20	60.6	719	17	AQ160089
c	30	20	60.6	735	9	AA706626
c	31	19.8	60.0	388	12	BE982337
c	32	19.8	60.0	472	10	BB829349
c	33	19.8	60.0	494	9	AL499781
c	34	19.8	60.0	495	10	BE634034
c	35	19.8	60.0	496	9	A1029512
c	36	19.8	60.0	504	17	BH809575
c	37	19.8	60.0	542	10	BE691497
c	38	19.8	60.0	600	17	AQ943317
c	39	19.8	60.0	615	14	BM942134
c	40	19.8	60.0	637	17	AQ950415
c	41	19.8	60.0	644	10	BB457668
c	42	19.8	60.0	744	17	AQ650190
c	43	19.6	59.4	572	10	AV742229
c	44	19.6	59.4	612	10	AV737807
c	45	19.6	59.4	682	13	BM397994

ALIGNMENTS

RESULT 1
LOCUS AV736417/c 615 bp mRNA linear EST 17-OCT-2000
DEFINITION AV736417 CB Homo sapiens CDNA clone CBND805 5', mRNA sequence.
ACCESSION AV736417
VERSION AV736417.1 GI:10853998
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Zhang, Q., Ye, M., Wu, X., Gu, J., Huang, Q., Zhou, J., Shen, Y., Han, Z.,
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Chen, S., Mao, M. and Chen, Z.
TITLE Homo sapiens CB library cDNA clones
JOURNAL Unpublished (2000)
COMMENT Shanghai Institute of Hematology, Rui-jin Hospital
197 Rui-jin II Road, Shanghai 200025, P. R. China.
Tel: 86-21-64740490
Fax: 86-21-64743206
Email: mbs@shim.su.sh.cn
This clone is available at Shanghai Hematology Institute in
Shanghai.
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong.
Location/Qualifiers
1..615
/organism="Homo sapiens"

FEATURES
source

CDS			
DEFINITION	AQ989608	593 bp	DNA linear GSS 14-AUG-2000
ACCESSION	FIC00171		Photorhabdus luminescens strain M14 M13 library
VERSION	Photorhabdus luminescens genomic clone FIC00171, DNA sequence.		
KEYWORDS	AQ989608.1	GI:9648202	GSS.
SOURCE	Photorhabdus luminescens		

ORGANISM	Photorhabdus luminescens Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Photorhabdus.					
REFERENCE	1 (bases 1 to 593)					
AUTHORS	firench-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T., Daborin,P.J., Bowen,D. and Blattner,F.R. A genomic sample sequence of the entomopathogenic bacterium Photorhabdus luminescens W14: potential implications for virulence Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000).					
JOURNAL MEDLINE	20378633 Contact: firench-Constant RH Department of Biology and Biochemistry University of Bath South Building, Bath BA2 7AY, UK Tel: (44) 1225 826621 Fax: (44) 1225 826779. Email: bsr@bath.ac.uk This is one of 2,122 random reads from the M13 library. For annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see firench-Constant et al. 2000. Nucleic Acids Res.					
COMMENT	Seq primer: M13 Forward Class: Shotgun. Location/Qualifiers 1..593 /organism="Photorhabdus luminescens" /strain="W14" /db_xref="taxon:29488" /clone="PLG00171" /clone_1fb-"Photorhabdus luminescens strain W14 M13 library" /dev_stage="primary phase variant" /note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus."					
BASE COUNT	169 a 166 c 115 g 143 t					
ORIGIN						
Query Match	66.7%; Score 22; DB 17; Length 593;					
Best Local Similarity	83.3%; Pred. No. 1.3e+02;					
Matches	25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;					
Dn	4 ACATGAGTGGCAAGCGCTGTGATACCG 33 405 ACTTGACTGCCAAAGCGTGTGATTACG 376					
RESULT 5						
AM804244	118 bp mRNA linear EST 16-MAY-2000					
LOCUS	AM804244					
DEFINITION	PH3-U00089-170300-002-b07 U00089_Homo sapiens CDNA, mRNA sequence.					
ACCESSION	AM804244					
VERSION	AM804244.1 GI:7856114					
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 118) Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Brlones,M.R., Nagai,M.A., da Silva,W.F., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.J., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.					
AUTHORS	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)					
TITLE	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 105, 4 andar, 01509-010, Sao Paulo-SP, Brazil.					
JOURNAL MEDLINE						
COMMENT						

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=6t2-PM3-UM0089-170>)
300-002-B07&t3=2000-03-17&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 7
High quality sequence stop: 76
Location/Qualifiers
1. .118
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_11b="UM0089"
/dev_stage="Adult"
/note="Organ: uterus; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESSES PCR (U.S. Letters Patent application No. 1967, 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT
ORIGIN
29 a 22 c 39 g 28 t

Query Match 65.5%; Score 21.6; DB 10; Length 118;
Best Local Similarity 85.7%; Pred. No. 97;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 CACATGAGTGCAGCGCTCTGTGATGATA 30'
1. 11111111 11111 1111 111111
Db 23 CTCATGAGTAGCAGACGCTGTGATGATA 50

RESULT 6
AZ922863/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

AZ922863 149 bp DNA linear GSS 07-JUN-2002
SICOT5F03 Sorghum bicolor SICOT sorghum bicolor genomic, DNA
sequence.
AZ922863
AZ922863.1 GI:13433084
GSS.
Sorghum.
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1. (bases 1 to 149)
Peterson, D.G., Schulze, S.R., Sciarra, E.B., Lee, S.A., Bowers, J.E.,
Nagel, A., Jiang, N., Tibbits, D.C., Wessler, S.R. and Peterson, A.H.
Integration of Cot analysis, DNA cloning, and high-throughput
sequencing facilitates genome characterization and gene discovery
Genome Res. 12 (5), 795-807 (2002)
21992826
Contact: Peterson DG
Plant Genome Mapping Laboratory
University of Georgia
Room 162, Riverbend Research Bldg., 110 Riverbend Rd., Athens, GA
30602, USA
Tel: 706-583-0167
Fax: 706-583-0160
Email: dgpetarches.uga.edu
Class: Hydroxyapatite-fractionated DNA.
Location/Qualifiers
1. .149
/organism="Sorghum bicolor"
/cultivar="BRx623"
/db_xref="taxon:4558"
/clone_11b="Sorghum bicolor SICOT"
/tissue_type="leaves"
/dev_stage="seedling"
/note="Vector: pGEM-TA-Basy; A Cot analysis was performed

FEATURES
SOURCE

AV744190/c 609 bp mRNA linear EST 17-OCT-2000
 LOCUS AV744190 CB Homo sapiens CDNA clone CBUEE04 5', mRNA sequence.
 DEFINITION AV744190
 ACCESSION AV744190
 VERSION AV744190.1 GI:10861771
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 609)
 AUTHORS Zhang, Q., Ye, M., Wu, X., Gu, J., Huang, Q., Zhou, J., Shen, Y., Han, Z.,
 Chen, S., Mao, M. and Chen, Z.
 TITLE Homo sapiens CB library cDNA clones
 JOURNAL Unpublished (2000)
 COMMENT Contact: Zhu Chen
 Shanghai Institute of Hematology, Rui-jin Hospital
 197 Rui-jin II Road, Shanghai 200025, P. R. China
 Tel: 86-21-64740490
 Fax: 86-21-64743206
 Email: mshlems.stn.sh.cn
 This clone is available at Shanghai Hematology Institute in
 Shanghai.
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong.
 Location/Qualifiers
 1..609
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CBUEE04"
 /clone_1lb="CB"
 /tissue_type="cord blood"
 /cell_type="CD34+ hematopoietic stem/progenitor cell"
 /lab_host="BM25.8"
 /note="Vector: pBluescript; Site_1: EcoRI; The insert is
 cloned randomly with the EcoRI digestion"
 BASE COUNT 165 a 194 c 107 g 141 t 2 others
 ORIGIN
 Query Match 63.6%; Score 21; DB 10; Length 609;
 Best Local Similarity 82.8%; Pred. No. 3.4e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Oy 1 GCCACATGATGTCGACGCTGTGTGAT 29
 |||||||||
 Db 538 GCTCATGATGTCGACGCTGTGTGAT 510
 |||||||||
 RESULT 10 778 bp mRNA linear EST 20-OCT-2000
 BE784735/c
 LOCUS BE784735
 DEFINITION 601473642P1 NIH-MGC_68 Homo sapiens CDNA clone IMAGE:3876443 5',
 mRNA sequence.
 ACCESSION BE784735
 VERSION BE784735.1 GI:10205933
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 778)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: DCTD/DMP/gazdar
 CNA Library Prepared: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LHAM9637 row: d column: 12
 High quality sequence stop: 633.
 Location/Qualifiers
 1..778
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3876443"
 /clone_1lb="NIH-MGC_68"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Lung; Vector: pCMV-Sport6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.8 kb. Library constructed by Life
 Technologies."
 BASE COUNT 259 a 237 c 110 g 172 t
 ORIGIN
 Query Match 63.6%; Score 21; DB 12; Length 778;
 Best Local Similarity 82.8%; Pred. No. 3.8e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Oy 1 GCCACATGATGTCGACGCTGTGTGAT 29
 |||||||||
 Db 488 GTCATGATGTCGACGCTGTGTGAT 460
 |||||||||
 RESULT 11 538 bp mRNA linear EST 19-JUL-2001
 BI296816/c
 LOCUS BI296816
 DEFINITION UI-R-DKO-cfz-g-01-0-UI.s1 UI-R-DKO Rattus norvegicus CDNA clone
 UI-R-DKO-cfz-g-01-0-UI 3', mRNA sequence.
 ACCESSION BI296816
 VERSION BI296816.1 GI:14961619
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 538)
 AUTHORS Bonaldi, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: mscares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to identify it as a clone from the
 normalized rat placenta pool library cDNA library preparation: M.B.
 Soares lab Clone distribution: clones will be available through
 Research Genetics (www.resgen.com)
 Seg primer: M13 Forward
 POLYA=yes.
 Location/Qualifiers
 1..538
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-DKO-cfz-g-01-0-UI"
 /clone_1lb="UI-R-DKO"
 /dev_stage="ADULT"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DKO


```

AV737831/C      613 bp      mRNA      linear      EST 17-OCT-2000
LOCUS          AV737831 CB Homo sapiens CDNA clone CBMARH09 5', mRNA sequence.
DEFINITION     AV737831
ACCESSION      AV737831
VERSION        AV737831.1 GI:10855412
KEYWORDS
SOURCE
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
                EST.
                human.
REFERENCE      1 (bases 1 to 613)
AUTHORS       Zhang, Q., Ye, M., Wu, X., Gu, J., Huang, Q., Zhou, J., Shen, Y., Han, Z.,
                Chen, S., Mao, M. and Chen, Z.
TITLE         Homo sapiens CB library cDNA clones
COMMENT       Unpublished (2000)
CONTACT       Zhu Chen
               Shanghai Institute of Hematology, Rui-Jin Hospital
               197 Rui-Jin II Road, Shanghai 200025, P. R. China
               Tel: 86-21-64740490
               Fax: 86-21-64743206
               Email: mbs@shim.sln.sh.cn
               This clone is available at Shanghai Hematology Institute in
               Shanghai.
               Chinese National Human Genome Center at Shanghai
               351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong.
               Location/Qualifiers
               1..613
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="CBMARH09"
               /clone_1lb="CB"
               /tissue_type="cord blood"
               /cell_type="CD34+ hematopoietic stem/progenitor cell"
               /lab_host="BM25.8"
               /note="Vector: pBluescript; Site_1: EcoRI; The insert is
               cloned randomly with the EcoRI digestion"
BASE COUNT     168 a      194 c      102 g      146 t
ORIGIN
Query Match    62.4%; Score 20.6; DB 10; Length 613;
Best Local Similarity 85.2%; Pred. No. 5e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3 CACATGAGTGGCAGCGCTGTGTGAT 29
        | ||||| ||| ||| ||| ||| |||
        536 CTCATGAGTGGCAGCGCTGTGTGAT 510

DB
RESULT 15
AI525664/C      777 bp      mRNA      linear      EST 09-AUG-1999
LOCUS          PT1.3_04_D05.r tumor1 Homo sapiens CDNA 5', mRNA sequence.
DEFINITION     AI525664
ACCESSION      AI525664
VERSION        AI525664.1 GI:4439799
KEYWORDS
SOURCE
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
                1 (bases 1 to 777)
                Huang, G.M., Ng, W.L., Farkas, J., He, L., Liang, H.A., Gordon, D., Yu, J.
                and Hood, L.
                Prostate Cancer expression profiling by cDNA sequencing analysis
                Genomics 59 (2), 178-186 (1999)
                99339982
                Contact: Guyang Mathew Huang
                Leroy Hood
                University of Washington
                Department of Molecular Biotechnology, Box 357730, University of
                Washington, Seattle, WA 98195
                Tel: 5106280100
                Fax: 5106280108
                Email: huanggm@yahoo.com.
TITLE         Prostate Cancer expression profiling by cDNA sequencing analysis
JOURNAL        Genomics 59 (2), 178-186 (1999)
MEDLINE
COMMENT

```

```

FEATURES
source
Location/Qualifiers
1..777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="tumor1"
/note="Organ: Prostate; Vector: pBluescript; Directional
cDNA library was constructed using Lambda ZP II kit
(Stratagene). mRNA was extracted from a frozen prostate
tumor tissue (Mayo Clinics)."
BASE COUNT     205 a      189 c      151 g      146 t      86 others
ORIGIN
Query Match    62.4%; Score 20.6; DB 9; Length 777;
Best Local Similarity 85.2%; Pred. No. 5.6e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3 CACATGAGTGGCAGCGCTGTGTGAT 29
        | ||||| ||| ||| ||| ||| |||
        DB      324 CTCATGAGTGGCAGCGCTGTGTGAT 298

Search completed: July 11, 2003, 02:25:43
Job time : 600.333 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 54.8629 Seconds
(Without alignments)
1067.241 Million cell updates/sec

Title: US-10-087-631B-6

Perfect score: 26

Sequence: 1 gcagaagcgtctagccatgagcgctta 26

Scoring table: IDENTITY_NUC

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: N_Geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	26	18	AAT67193
2	26	100.0	26	19	AAV59058
3	26	100.0	26	22	AAH25413
4	26	100.0	53	24	AAI40117
5	26	100.0	73	24	AAI40115
6	26	100.0	77	22	AAI04490
7	26	100.0	153	21	AAZ57775
8	26	100.0	202	12	AAQ14083
9	26	100.0	202	12	AAQ14085

10	26	100.0	202	12	AAQ14086	HCV-N (1-200). He
11	26	100.0	202	12	AAQ14084	HCV-N (1-202). He
12	26	100.0	242	14	AAQ37774	Cloned HCV 5' non
13	26	100.0	244	19	AAV70449	HCV subtype 1a PCR
14	26	100.0	244	19	AAV70450	HCV subtype 1b PCR
15	26	100.0	244	19	AAV70451	HCV subtype 2c PCR
16	26	100.0	244	19	ABLA6059	Hepatitis C virus
17	26	100.0	244	24	ABLA6060	Hepatitis C virus
18	26	100.0	244	24	ABLA6061	Hepatitis C virus
19	26	100.0	260	21	AAZ57774	Hepatitis C virus
20	26	100.0	279	24	ABA01119	Hepatitis C virus
21	26	100.0	281	17	AAT29118	Hepatitis C virus
22	26	100.0	281	17	AAT29120	Hepatitis C virus
23	26	100.0	282	17	AAT29119	Hepatitis C virus
24	26	100.0	286	19	AAV70444	HCV subtype 1b tar
25	26	100.0	286	24	ABLA6054	Hepatitis C virus
26	26	100.0	289	19	AAV70443	HCV consensus sequ
27	26	100.0	289	19	AAV70445	HCV subtype 2c tar
28	26	100.0	289	24	ABLA6053	Hepatitis C virus
29	26	100.0	289	24	ABLA6055	Hepatitis C virus
30	26	100.0	305	18	AAH70088	HCV amplification
31	26	100.0	305	24	ABN79959	Hepatitis C virus
32	26	100.0	305	24	ABN79970	Hepatitis C virus
33	26	100.0	305	24	ABN79971	Hepatitis C virus
34	26	100.0	305	24	ABN79972	Hepatitis C virus
35	26	100.0	306	15	AAO67079	Hepatitis C virus
36	26	100.0	308	21	AAH75294	Novel hepatitis C
37	26	100.0	312	21	AAZ36198	Adapted HCV 5' non
38	26	100.0	314	21	AAZ36197	Adapted HCV 5' non
39	26	100.0	323	24	ABK70882	HCV genome 5'UTR s
40	26	100.0	323	24	ABK70883	HCV genome 5'UTR s
41	26	100.0	326	24	ABK70880	HCV genome 5'UTR s
42	26	100.0	327	21	AAZ36199	Adapted HCV 5' non
43	26	100.0	327	21	ABK70884	HCV genome 5'UTR s
44	26	100.0	328	18	AAT77074	Hepatitis C virus
45	26	100.0	328	24	ABLA6273	Hepatitis C virus

ALIGNMENTS

RESULT 1	
AAT67193	standard; DNA; 26 BP.
AC	AAT67193:
DT	13-FEB-1998 (first entry)
DE	Hepatitis C virus (HCV) RNA amplification primer ST280A.
KW	Hepatitis C virus; HCV; ST280A; reverse transcription PCR; RT-PCR;
KW	PCR primer; ss.
OS	Synthetic.
PN	EP776981-A2.
PD	04-JUN-1997.
PF	21-NOV-1996; 96EP-0118704.
PR	29-NOV-1995; 95US-0007739.
PA	(HOFF) HOFFMANN LA ROCHE & CO AG F.
PI	Tsang SY;
PT	WPI: 1997-291296/27.
DR	Oligonucleotide primers for hepatitis C virus RNA amplification
PT	by polymerase chain reaction
XX	

PS Claim 1; Page 11; 16pp; English.

CC This upstream primer, ST280A is used in the amplification of the
CC Hepatitis C virus (HCV) RNA by reverse transcription PCR. This is used
CC to amplify a 250 base pair product from the 5' untranslated region of
CC the HCV genome. This can be used to detect HCV in a sample with increased
CC sensitivity. Amplification of HCV nucleic acid using this primer is up to
CC 100 times more efficient than amplification with prior art primers.

SQ Sequence 26 BP; 7 A; 6 C; 8 G; 5 T; 0 other;

Query Match 100.0%; Score 26; DB 19; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTCTAGCCATGCGCTTA 26
1 GCAGAAAGCGCTCTAGCCATGCGCTTA 26

RESULT 2

ID AAV59058 standard; DNA; 26 BP.

XX AAV59058;

DT 07-JAN-1999 (first entry)

DE Primer ST280A for HCV fragment.

KW PCR primer; HCV; nucleic acid amplification; ss.

OS Synthetic.

XX Human cytomegalovirus.

FT modified base 26

FT /tag- a Location/Qualifiers
FT /note- "optionally benzylated, methylated, or
FT nitrobenzylated"

XX BP866071-A2.

PD 23-SEP-1998.

PF 12-MAR-1998; 98BP-0104461.

PR 20-MAR-1997; 97US-0041127.

XX (HOFF) HOFFMANN LA ROCHE & CO AG F.

XX Will SG, Young KKY.

DR WPI; 1998-482929/42.

PT Oligo-nucleotide(s) containing N-substituted nucleotide - useful as
PT primers for nucleic acid amplification

PS Example 6; Page 16; 38pp; English.

CC This sequence represents a primer for a fragment of HCV, and is an
CC example of an oligonucleotide of the invention. The oligonucleotides of
CC the invention are of the formula 5'-S1-Nu-3' or 5'-S1-Nu-S2-3', where
CC S1 is a sequence of 5-50 nucleotides; S2 is a sequence of 1-3
CC nucleotides; and Nu is a nucleotide with a purine or pyrimidine base
CC having an exocyclic amino group substituted by a phenyl, phenoxy or optionally
CC 1-10C-alkyl, alkoxy, optionally substituted phenyl, phenoxy or optionally
CC substituted naphthyl. The oligonucleotides are useful as primers for
CC nucleic acid amplification, preferably by polymerase chain reaction. Use
CC of the modified primers reduces non-specific amplification, especially
CC primer dimer formation, with a concomitant increase in the yield of the
CC intended target.

SQ Sequence 26 BP; 7 A; 6 C; 8 G; 5 T; 0 other;

Query Match 100.0%; Score 26; DB 19; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTCTAGCCATGCGCTTA 26
1 GCAGAAAGCGCTCTAGCCATGCGCTTA 26

Db 1 GCAGAAAGCGCTCTAGCCATGCGCTTA 26

RESULT 3

ID AAH25413 standard; DNA; 26 BP.

XX AAH25413;

DT 22-AUG-2001 (first entry)

DE Forward PCR primer used to amplify a HCV DNA fragment.

KW Magnetic glass particle; nucleic acid purification; PCR primer; ss.

OS Hepatitis C virus.

FT Key Location/Qualifiers

FT modified base 26

FT /tag- a /note- "derivatisation with a p-(t-butyl)benzyl-residue"

XX WO200137291-A1.

PD 25-MAY-2001.

PF 17-NOV-2000; 2000MO-EP11459.

PR 17-NOV-1999; 98BP-0122853.

XX 12-MAY-2000; 2000EP-0110165.

XX (HOFF) ROCHE DIAGNOSTICS GMBH.

XX Weindel K, Riedling M, Geiger A;

DR WPI; 2001-381247/40.

PT Novel composition of magnetic glass particles for purification of DNA
PT or RNA in automated processes

PS Example 7; Page 98; 105pp; English.

CC The specification describes a composition of magnetic glass particles,
CC which contain at least one magnetic object with a mean diameter between
CC 5-500 nm. The composition is useful for the purification of nucleic
CC acids. The composition can be used to process large quantities of
CC nucleic acid samples, because it does not involve the particles being
CC centrifuged or the fluids being drawn through glass fiber filters.
CC PCR primers AAH25413-14 were used to amplify HCV DNA fragments. The
CC amplified fragment can be purified using the method of the invention.

SQ Sequence 26 BP; 7 A; 6 C; 8 G; 5 T; 0 other;

Query Match 100.0%; Score 26; DB 22; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTCTAGCCATGCGCTTA 26
1 GCAGAAAGCGCTCTAGCCATGCGCTTA 26

RESULT 4

ID AAL40117 standard; DNA; 53 BP.

[illegible]

```
/*tag= e
/bound_mo
```

18-DEC-2000; 2000WO-GB04862.
16-DEC-1999; 99GB-0029920.
22-DEC-1999; 99US-0171804.
(RIBO-) RIBOTANGENTS LTD.
Karn J, Walker S;
WPI; 2001-465050/50.
Nucleotide sequences derived from Hepatitis C virus, useful for
identifying candidate antiviral compounds -
Disclosure; Fig 5E; 48pp; English.

domain in 18S rRNA probe used in a RNA electrophoretic gel mobility shift assay (EMSA). The present sequence is described in an invention relating to a novel compound comprising nucleotide sequences capable of annealing to a region of HCV which is essential for binding of eIF3 (eukaryotic initiation factor 3). The invention particularly relates to a sub-region of the HCV 5'-UTR referred to as the minimal internal ribosome entry site (mIRES) which can be used to identify internal inhibitors of HCV translation initiation. The compounds of the invention may be used to screen for potential HCV antiviral compounds. Assays based on the mIRES enable potential antivirals to be screened in a cheaper and easier way. It allows rapid assaying with a small volume of material and are suitable to parallel processing.

```

    100.0%   Score 26; DB 22; Length 77;
    at Local Similarity 80.8%; Pred. No. 0.0027;
    ches 21; Conservative 15; Mismatches 0; Indels 0; Gaps 0

```

1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
26 GCAGAAAGCGUCUAGCCAUGGCGUUA 51

777

AAZ57775 standard; DNA; 155 BP

05-APR-2000 (first entry)

inhibitor oligonucleotide #41

anti-inflammatory; translation inhibition; HCV infection; virucide.

Hepatitis C virus

IS6001990-A

4-DEC-1999

7-JUN-1995; 95US-0474700

0-MAY-1994; 94US-0240382

(GEHO) GEN HOSPITAL CORP

Moradpour D, Wands JR, Wakita T;

WPI; 2000-104900/09

amplisense oligonucleotide to Hepatitis C virus RNA, useful for treating Hepatitis C virus infections -

Claim 30; Column 33; 31pp; English

Hepatitis sequence is an antisense oligonucleotide that hybridises to Hepatitis C virus (HCV) RNA, under physiological conditions. The invention relates to HCV antisense oligonucleotides, and also for a vector comprising a nucleotide sequence which is transcribed in an animal cell to generate an antisense oligonucleotide. The oligonucleotides have a 5' triphosphate, a 3' hydroxyl, a 5' triphosphate, and are useful for treating HCV infection by inhibiting translational activity, and HCV RNA. Hepatitis C virus is a positive strand RNA virus, and is the major causative agent of post-transfusion hepatitis. Persistent HCV infection can lead to chronic hepatitis, cirrhosis, and hepatocellular carcinoma.

sequence 133 BP; 29 A; 43 C; 51 G; 32 T; 0 other;

Local Match	100.0%;	Score 26;	DB 21;	Length 155;
Local Similarity	100.0%;	Pred. No. 0.003;		
Conservative	0;	Mismatches 0;	T+G-T	
26;				

1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
88 GCAGAAAGCGTCTAGCCATGGCGTTA 63

83
AQ14083 standard; DNA; 202 BP

AQ14083;

3-JAN-1992 (first entry)

CV-1 (1-202)

ptopoe; PCR; diagnosis; ss.

hepatitis C virus

09114779-A

3 - OCT - 1991

8-MAR-1991; 91WO-JP00405

90JP-0305795

-JUN-1990; 90JP-0154230.

433

THE

Т. 1

I; 1991-310579/42

hepatitis C virus infection via polymerase chain reaction
nucleotide sequences encoding HCV epitope(s) - for diagnosis
hepatitis C virus infection via polymerase chain reaction
closure; Fig 1(1); 79pp; Japanese.

code epitopes from structural, non-structural and 5' untranslated regions of hepatitis C virus. The sequences are used for accurate diagnosis and typing of HCV infection, using PCR.

CC amplification techniques.

SQ Sequence 202 BP; 39 A; 61 C; 61 G; 41 T; 0 other;

Query Match 100.0%; Score 26; DB 12; Length 202;

Best Local Similarity 100.0%; Pred. No. 0.0032;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
 |||||||
 DB 25 GCAGAAAGCGTCTAGCCATGGCGTTA 50

RESULT 9

AAQ14085 ID AAQ14085 standard; DNA; 202 BP.

AC AAQ14085;

DT 03-JAN-1992 (first entry)

DE HCV-KU (1-202).

KM Epitope; PCR; diagnosis; ss.

OS Hepatitis C virus.

PN WO9114779-A.

PD 03-OCT-1991.

PF 28-MAR-1991; 91WO-JP00405.

PR 09-NOV-1990; 90JP-0305795.

PR 28-MAR-1990; 90JP-0080185.

PR 13-JUN-1990; 90JP-0154230.

PR 14-JUN-1990; 90JP-0153979.

PA (MITK) MITSUI TOATSU CHEM INC.

PI Takada T, Enomoto N, Date T, Nakao T;

DR WPI; 1991-310579/42.

PS Disclosure; Fig 1(I); 79pp; Japanese.

CC The nucleotide sequences represented in AAQ14076-86 and AAQ14767-71
 CC encode epitopes from structural, non-structural and 5' untranslated
 CC domains of hepatitis C virus. The sequences are used for accurate
 CC and simple diagnosis and typing of HCV infection, using PCR
 CC amplification techniques.

SQ Sequence 202 BP; 40 A; 61 C; 58 G; 43 T; 0 other;

Query Match 100.0%; Score 26; DB 12; Length 202;

Best Local Similarity 100.0%; Pred. No. 0.0032;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
 |||||||
 DB 25 GCAGAAAGCGTCTAGCCATGGCGTTA 50

RESULT 10

AAQ14086 ID AAQ14086 standard; DNA; 202 BP.

AC AAQ14086;

DT 03-JAN-1992 (first entry)

XX HCV-N (1-200).

XX Epitope; PCR; diagnosis; ss.

KW Hepatitis C virus.

PN WO9114779-A.

PD 03-OCT-1991.

PF 28-MAR-1991; 91WO-JP00405.

PR 09-NOV-1990; 90JP-0305795.

PR 28-MAR-1990; 90JP-0080185.

PR 13-JUN-1990; 90JP-0154230.

PR 14-JUN-1990; 90JP-0153979.

PA (MITK) MITSUI TOATSU CHEM INC.

PI Takada T, Enomoto N, Date T, Nakao T;

DR WPI; 1991-310579/42.

PS Disclosure; Fig 1(I); 79pp; Japanese.

CC The nucleotide sequences represented in AAQ14076-86 and AAQ14767-71
 CC encode epitopes from structural, non-structural and 5' untranslated
 CC domains of hepatitis C virus. The sequences are used for accurate
 CC and simple diagnosis and typing of HCV infection, using PCR
 CC amplification techniques.

SQ Sequence 202 BP; 42 A; 60 C; 56 G; 44 T; 0 other;

Query Match 100.0%; Score 26; DB 12; Length 202;

Best Local Similarity 100.0%; Pred. No. 0.0032;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
 |||||||
 DB 25 GCAGAAAGCGTCTAGCCATGGCGTTA 50

RESULT 11

AAQ14084 ID AAQ14084 standard; DNA; 202 BP.

AC AAQ14084;

DT 03-JAN-1992 (first entry)

DE HCV-I (1-202).

KM Epitope; PCR; diagnosis; ss.

OS Hepatitis C virus.

PN WO9114779-A.

PD 03-OCT-1991.

PF 28-MAR-1991; 91WO-JP00405.

PR 09-NOV-1990; 90JP-0305795.

PR 28-MAR-1990; 90JP-0080185.

PR 13-JUN-1990; 90JP-0154230.

PR 14-JUN-1990; 90JP-0153979.

PA (MITK) MITSUI TOATSU CHEM INC.

PI Takada T, Enomoto N, Date T, Nakao T;
 XX
 DR WPI: 1991-310579/42.
 XX
 PT New nucleotide sequences encoding HCV epitope(s) - for diagnosis
 PT of hepatitis C virus infection via polymerase chain reaction
 XX
 PS Disclosure: Fig 1(1): 79pp; Japanese.
 CC The nucleotide sequences represented in AA014076-86 and AA014767-71
 CC encode epitopes from structural, non-structural and 5' untranslated
 CC domain of hepatitis C virus. The sequences are used for accurate
 CC and simple diagnosis and typing of HCV infection, using PCR
 CC amplification techniques.
 XX
 SQ Sequence 202 BP; 41 A; 60 C; 60 G; 41 T; 0 other;
 Query Match 100.0%; Score 26; DB 12; Length 202;
 Best Local Similarity 100.0%; Pred. No. 0.0032;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GCAGAAAGCGCTAGCCATGGCGTTA 26
 DB 25 GCAGAAAGCGCTAGCCATGGCGTTA 50
 RESULT 12
 AA037774
 ID AA037774 standard; cDNA; 242 BP.
 AC AA037774;
 XX
 DT 30-JUN-1993 (first entry)
 XX
 DE Cloned HCV 5' non coding region from pGHCV1A.
 XX
 KW Hepatitis C virus; probe; hepatocellular necrosis; hepatocellular;
 KW carcinoma; diagnosis; therapy; ss.
 OS Hepatitis C virus.
 XX
 PN EP531974-A.
 PD 17-MAR-1993.
 XX
 PF 09-SEP-1992; 92EP-0115426.
 XX
 PR 12-SEP-1991; 91US-0758862.
 XX
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 XX
 PI Hu K, Vierling JM;
 DR WPI: 1993-087007/11.
 XX
 PT Detection of hepatitis C virus (HCV) RNA - using nucleic acid
 PT probes derived from the 5'-non-coding region of the HCV genome
 XX
 PS Claim 1; Fig 4; 26pp; English.
 CC To obtain HCV cDNA nucleotide sequences from the 5' non-coding
 CC region a pair of oligonucleotides based on the reported sequence of
 CC HCV-J1 were used as primers for HCV PCR. HCV RNA was isolated from
 CC serum of a putatively infected individual. RNA reverse
 CC transcription PCR was performed and a specific PCR prod. identified.
 CC The prod. was used to transform E. coli DH5 alpha to obtain pGHCV1A
 CC contg. a 242 bp insertion from the HCV 5' non-coding region. This
 CC probe is highly specific and sensitive for HCV RNA. The probe can
 CC be used to quantitatively detect the amt. of HCV in samples, to
 CC analyse the molecular forms of HCV RNA during evolution of the
 CC disease, to localise HCV in hepatic and/or extrahepatic tissues
 CC and to study the relationship between HCV infection, hepatocellular
 CC necrosis and hepatocellular carcinoma. The probe can be used to

CC diagnose HCV infection, to prepare blood free of HCV and to monitor
 CC anti-HCV therapy.
 XX
 SQ Sequence 242 BP; 51 A; 74 C; 67 G; 50 T; 0 other;
 Query Match 100.0%; Score 26; DB 14; Length 242;
 Best Local Similarity 100.0%; Pred. No. 0.0033;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GCAGAAAGCGCTAGCCATGGCGTTA 26
 DB 45 GCAGAAAGCGCTAGCCATGGCGTTA 70
 RESULT 13
 AA070449
 ID AA070449 standard; DNA; 244 BP.
 AC AA070449;
 XX
 DT 08-APR-1999 (first entry)
 XX
 DE HCV subtype 1a PCR fragment.
 XX
 KW Nucleic acid detection; nucleic acid characterisation; hybridisation;
 KW infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.
 OS Hepatitis C virus.
 XX
 PN WC9850403-A1.
 PD 12-NOV-1998.
 XX
 PF 05-MAY-1998; 98WO-US03194.
 XX
 PR 03-MAR-1998; 98US-0034205.
 PR 05-MAY-1997; 97US-0851588.
 PR 19-SEP-1997; 97US-0934097.
 XX
 PA (THIR-) THIRD WAVE TECHNOLOGIES INC.
 XX
 PI Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L;
 PI Lyamatchev VI, Neri BP, Prudent JR;
 DR WPI: 1998-610317/51.
 XX
 PT Detection and characterisation of nucleic acid sequences - by mixing
 PT a folded target and one or more probes to form a probe/folded target
 PT complex and detecting and characterising the complexes
 XX
 PS Example 3; Page 169; 279pp; English.
 CC The invention relates to methods and compositions of detection and
 CC characterisation of nucleic acid sequences and sequence changes. One
 CC method of detection and characterisation comprises: (a) providing: (i) a
 CC folded target having a DNA sequence comprising at least 1 double
 CC stranded region and at least 1 single stranded region; and (ii) at least
 CC 1 probe complementary to at least a portion of the folded target; and
 CC (b) mixing the target and probes so that the probe hybridises to form a
 CC of structure formation in nucleic acid targets; for analysing folded
 CC nucleic acids targets; and for analysis of nucleic acid structures. The
 CC method can be used for the detection and characterisation of nucleic
 CC acid sequences indicative of an infection, the presence of pathogenic nucleic acid
 CC sequences associated with disease and cancers, and the
 CC identification of the source of nucleic acids found in forensic samples,
 CC as well as in paternity determinations. The methods allow simultaneous
 CC analysis of both strands (e.g. the sense and antisense strands) and are
 CC ideal for high-level multiplexing. The products produced are amenable to
 CC qualitative, quantitative and positional analysis. The methods may be
 CC performed in solution or in the solid phase (e.g. on a solid support).
 CC The methods are powerful in that they allow for analysis of longer

CC fragments of nucleic acid than current methodologies. Sequences
CC AAV70449-52 represent a hepatitis C virus (HCV) subtype sequences
CC produced by PCR. These PCR products can be used in hybridisation analysis
CC using multiple capture probes for HCV genotyping.

CC Sequence 244 BP; 46 A; 67 C; 80 G; 51 T; 0 other;

SO Query Match 100.0%; Score 26; DB 19; Length 244;

Best Local Similarity 100.0%; Pred. No. 0.0033; Mismatches 0; Indels 0; Gaps 0;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCCATGCGCTTA 26

1 GCAGAAAGCGCTAGCCATGCGCTTA 26

Db

AAV70450 standard; DNA; 244 BP.

AAV70450;

08-APR-1999 (first entry)

HCV subtype 1b PCR fragment.

Nucleic acid detection; nucleic acid characterisation; hybridisation;

Infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.

Hepatitis C virus.

MO9850403-A1.

12-NOV-1998.

05-MAY-1998; 98MO-US03194.

03-MAR-1998; 98US-0034205.

05-MAY-1997; 97US-0851588.

19-SEP-1997; 97US-0934097.

(THIR-) THIRD WAVE TECHNOLOGIES INC.

Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L;

Lyamichev VI, Neri BP, Prudent JR;

WPI; 1998-610317/51.

Detection and characterisation of nucleic acid sequences - by mixing

a folded target and one or more probes to form a probe/folded target

complex and detecting and characterising the complexes

Example 3; Page 169; 279pp; English.

The invention relates to methods and compositions of detection and

characterisation of nucleic acid sequences and sequence changes. One

method of detection and characterisation comprises: (a) providing: (i) a

folded target having a DNA sequence comprising at least 1 double

stranded region and at least 1 single stranded region; and (ii) at least

1 probe complementary to at least a portion of the folded target; and

(b) mixing the target and probes so that the probe hybridises to form a

probe/folded target complex. Also provided are methods for determination

of structure formation in nucleic acid targets; for analysing folded

nucleic acids targets; and for analysis of nucleic acid structures. The

methods can be used for the detection and characterisation of nucleic

acid sequences to detect the presence of pathogenic nucleic acid

sequences indicative of an infection, the presence of variants or alleles

of mammalian genes associated with disease and cancers, and the

identification of the source of nucleic acids found in forensic samples,

as well as in paternity determinations. The methods allow simultaneous

analysis of both strands (e.g. the sense and antisense strands) and are

ideal for high-level multiplexing. The products produced are amenable to

qualitative, quantitative and positional analysis. The methods may be

CC performed in solution or in the solid phase (e.g. on a solid support).

CC The methods are powerful in that they allow for analysis of longer

CC fragments of nucleic acid than current methodologies. Sequences

CC AAV70449-52 represent a hepatitis C virus (HCV) subtype sequences

CC produced by PCR. These PCR products can be used in hybridisation analysis

CC using multiple capture probes for HCV genotyping.

CC Sequence 244 BP; 44 A; 67 C; 81 G; 52 T; 0 other;

SO Query Match 100.0%; Score 26; DB 19; Length 244;

Best Local Similarity 100.0%; Pred. No. 0.0033; Mismatches 0; Indels 0; Gaps 0;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCCATGCGCTTA 26

1 GCAGAAAGCGCTAGCCATGCGCTTA 26

Db

AAV70451 standard; DNA; 244 BP.

AAV70451;

08-APR-1999 (first entry)

HCV subtype 2c PCR fragment.

Nucleic acid detection; nucleic acid characterisation; hybridisation;

Infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.

Hepatitis C virus.

MO9850403-A1.

12-NOV-1998.

05-MAY-1998; 98MO-US03194.

03-MAR-1998; 98US-0034205.

05-MAY-1997; 97US-0851588.

19-SEP-1997; 97US-0934097.

(THIR-) THIRD WAVE TECHNOLOGIES INC.

Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L;

Lyamichev VI, Neri BP, Prudent JR;

WPI; 1998-610317/51.

Detection and characterisation of nucleic acid sequences - by mixing

a folded target and one or more probes to form a probe/folded target

complex and detecting and characterising the complexes

Example 3; Page 169; 279pp; English.

The invention relates to methods and compositions of detection and

characterisation of nucleic acid sequences and sequence changes. One

method of detection and characterisation comprises: (a) providing: (i) a

folded target having a DNA sequence comprising at least 1 double

stranded region and at least 1 single stranded region; and (ii) at least

1 probe complementary to at least a portion of the folded target; and

(b) mixing the target and probes so that the probe hybridises to form a

probe/folded target complex. Also provided are methods for determination

of structure formation in nucleic acid targets; for analysing folded

nucleic acids targets; and for analysis of nucleic acid structures. The

methods can be used for the detection and characterisation of nucleic

acid sequences to detect the presence of pathogenic nucleic acid

sequences indicative of an infection, the presence of variants or alleles

of mammalian genes associated with disease and cancers, and the

identification of the source of nucleic acids found in forensic samples,

as well as in paternity determinations. The methods allow simultaneous

analysis of both strands (e.g. the sense and antisense strands) and are

CC Ideal for high-level multiplexing. The products produced are amenable to
 CC qualitative, quantitative and positional analysis. The methods may be
 CC performed in solution or in the solid phase (e.g. on a solid support).
 CC The methods are powerful in that they allow for analysis of longer
 CC fragments of nucleic acid than current methodologies. Sequences
 CC AAV70449-52 represent a hepatitis C virus (HCV) subtype sequences
 CC produced by PCR. These PCR products can be used in hybridisation analysis
 CC using multiple capture probes for HCV genotyping.

XX .sequence 244 BP; 46 A; 68 C; 78 G; 52 T; 0 other;

Query Match 100.0%; Score 26; DB 19; Length 244;

Best Local Similarity 100.0%; Pred. No. 0.0033;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCCATGCGCTTA 26
 |||||
 DB 1 GCAGAAAGCGCTAGCCATGCGCTTA 26

Search completed: July 10, 2003, 19:52:41
 Job time : 56.8629 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 ; Search time 11.627 Seconds
(without alignments)
685.785 Million cell updates/sec

Title: US-10-087-631B-6

Sequence: 1 gcagaaagcgtctagccatgcgcgtta 26

Scoring table: IDENTITY_MUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PTC05.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	100.0	26	US-08-738-928-1	Sequence 1, Appl
2	26	100.0	26	US-09-039-866-3	Sequence 3, Appl
3	26	100.0	155	US-08-474-700B-41	Sequence 41, Appl
4	26	100.0	194	US-08-634-797-44	Sequence 44, Appl
5	26	100.0	194	US-08-634-797-45	Sequence 45, Appl
6	26	100.0	194	US-08-634-797-46	Sequence 46, Appl
7	26	100.0	194	US-08-634-797-47	Sequence 47, Appl
8	26	100.0	194	US-08-634-797-49	Sequence 49, Appl
9	26	100.0	194	US-08-634-797-50	Sequence 50, Appl
10	26	100.0	194	US-08-634-797-51	Sequence 51, Appl
11	26	100.0	194	US-08-634-797-52	Sequence 52, Appl
12	26	100.0	244	US-08-335-595-1	Sequence 1, Appl
13	26	100.0	244	US-09-034-205-26	Sequence 26, Appl
14	26	100.0	244	US-09-034-205-27	Sequence 27, Appl
15	26	100.0	244	US-09-034-205-28	Sequence 28, Appl
16	26	100.0	244	US-08-934-097A-26	Sequence 26, Appl
17	26	100.0	244	US-08-934-097A-27	Sequence 27, Appl
18	26	100.0	244	US-08-934-097A-28	Sequence 28, Appl
19	26	100.0	244	US-08-851-588-26	Sequence 26, Appl
20	26	100.0	244	US-08-851-588-27	Sequence 27, Appl
21	26	100.0	244	US-08-851-588-28	Sequence 28, Appl
22	26	100.0	244	US-09-677-218B-26	Sequence 26, Appl
23	26	100.0	244	US-09-677-218B-27	Sequence 27, Appl
24	26	100.0	244	US-09-677-218B-28	Sequence 28, Appl
25	26	100.0	244	US-09-677-192-26	Sequence 26, Appl
26	26	100.0	244	US-09-677-192-27	Sequence 27, Appl
27	26	100.0	244	US-09-677-192-28	Sequence 28, Appl

c	28	26	100.0	260	3	US-08-474-700B-40	Sequence 40, Appl
c	29	26	100.0	281	2	US-08-757-653-123	Sequence 123, App
c	30	26	100.0	281	2	US-08-757-653-125	Sequence 125, App
c	31	26	100.0	281	2	US-08-757-653-128	Sequence 128, App
c	32	26	100.0	281	2	US-08-757-653-129	Sequence 129, App
c	33	26	100.0	281	2	US-08-757-653-131	Sequence 131, App
c	34	26	100.0	281	4	US-08-520-946-123	Sequence 123, App
c	35	26	100.0	281	4	US-08-520-946-125	Sequence 125, App
c	36	26	100.0	281	4	US-08-520-946-128	Sequence 128, App
c	37	26	100.0	281	4	US-08-520-946-129	Sequence 129, App
c	38	26	100.0	281	4	US-08-520-946-131	Sequence 131, App
c	39	26	100.0	282	2	US-08-757-653-134	Sequence 134, App
c	40	26	100.0	282	2	US-08-757-653-130	Sequence 130, App
c	41	26	100.0	282	4	US-08-520-946-124	Sequence 124, App
c	42	26	100.0	282	4	US-08-520-946-130	Sequence 130, App
c	43	26	100.0	286	4	US-09-034-205-21	Sequence 21, App
c	44	26	100.0	286	4	US-08-934-097A-21	Sequence 21, App
c	45	26	100.0	286	4	US-08-851-588-21	Sequence 21, App

ALIGNMENTS

RESULT 1
US-08-738-928-1
Sequence 1, Application US/08738928
Patent No. 5837442
GENERAL INFORMATION:
APPLICANT: Tsang, Sue Y.
TITLE OF INVENTION: Oligonucleotide primers for amplifying
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/738,928
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Petry, Douglas A.
REGISTRATION NUMBER: 35,321
REFERENCE/DOCKET NUMBER: 9263
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2974
TELEFAX: (510) 814-2974
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 26 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-738-928-1
Query Match 100.0%; Score 26; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.00067;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCCTTACCCATGCGGCTTA 26
DB 1 GCAGAAAGCCTTACCCATGCGGCTTA 26

RESULT 2
US-09-039-866-3
Sequence 3, Application US/09039866
Patent No. 6001611
GENERAL INFORMATION:
APPLICANT: Will, Stephen G.
TITLE OF INVENTION: MODIFIED NUCLEIC ACID AMPLIFICATION
TITLE OF INVENTION: PRIMERS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Roche Molecular Systems
STREET: 1080 U.S. Highway 202
City: Branchburg
STATE: New Jersey
COUNTRY: United States
ZIP: 08876
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/039,866
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Petry, Douglas A.
REGISTRATION NUMBER: 35,321
REFERENCE/DOCKET NUMBER: 1023P
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 26 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-039-866-3
Query Match 100.0%; Score 26; DB 3; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.00067;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GCAGAAAGCGCTAGCCATGGCGTTA 26
DB 1 GCAGAAAGCGCTAGCCATGGCGTTA 26
RESULT 3
US-08-474-700B-41/c
Sequence 41, Application US/08474700B
Patent No. 6001990
GENERAL INFORMATION:
APPLICANT: Wands, Jack
APPLICANT: Wakisaka, Takaji
TITLE OF INVENTION: ANTISENSE INHIBITION OF HEPATITIS C
TITLE OF INVENTION: VIRUS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
City: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB.
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,700B
FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/240,382
FILING DATE: 10 May 1994
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00786/279001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 155 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-474-700B-41
Query Match 100.0%; Score 26; DB 3; Length 155;
Best Local Similarity 100.0%; Pred. No. 0.00084;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GCAGAAAGCGCTAGCCATGGCGTTA 26
DB 88 GCAGAAAGCGCTAGCCATGGCGTTA 63
RESULT 4
US-08-634-797-44
Sequence 44, Application US/08634797
Patent No. 5831759
GENERAL INFORMATION:
APPLICANT: WEINER, AMY J.
TITLE OF INVENTION: HETEROIDPLEX TRACKING ASSAY (HTA) FOR
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street - R440
City: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,797
FILING DATE: 19-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hardin, Alisha A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 1226.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-3274
TELEFAX: (510) 655-3542
TELEX: N/A
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-634-797-44
Query Match 100.0%; Score 26; DB 2; Length 194;
Best Local Similarity 100.0%; Pred. No. 0.00086;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
DB 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26

RESULT 5

US-08-634-797-45
Sequence 45, Application US/08634797
Patent No. 5851759

GENERAL INFORMATION:
APPLICANT: WEINER, AMY J.
TITLE OF INVENTION: HETEROIDUPLEX TRACKING ASSAY (HTA) FOR
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street - R440
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/634,797
APPLICATION NUMBER: US/08/634,797
FILING DATE: 19-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hardin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 1226.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-3274
TELEFAX: (510) 655-3542
TELEX: N/A
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-634-797-45

Query Match 100.0%; Score 26; DB 2; Length 194;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
DB 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26

RESULT 6

US-08-634-797-46
Sequence 46, Application US/08634797
Patent No. 5851759

GENERAL INFORMATION:
APPLICANT: WEINER, AMY J.
TITLE OF INVENTION: HETEROIDUPLEX TRACKING ASSAY (HTA) FOR
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street - R440
CITY: Emeryville
STATE: California
COUNTRY: USA

ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/634,797
APPLICATION NUMBER: US/08/634,797
FILING DATE: 19-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hardin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 1226.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-3274
TELEFAX: (510) 655-3542
TELEX: N/A
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-634-797-46

Query Match 100.0%; Score 26; DB 2; Length 194;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
DB 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26

RESULT 7

US-08-634-797-47
Sequence 47, Application US/08634797
Patent No. 5851759

GENERAL INFORMATION:
APPLICANT: WEINER, AMY J.
TITLE OF INVENTION: HETEROIDUPLEX TRACKING ASSAY (HTA) FOR
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street - R440
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/634,797
APPLICATION NUMBER: US/08/634,797
FILING DATE: 19-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hardin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 1226.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-3274
TELEFAX: (510) 655-3542
TELEX: N/A
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-634-797-47

Query Match
Best Local Similarity 100.0%; Score 26; DB 2; Length 194;
Matches 26; Conservative 0; Pred. No. 0.00086;
Mismatched 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCATGCGCTTA 26
Db 1 GCAGAAAGCGCTAGCATGCGCTTA 26

RESULT 8
US-08-634-797-49

Sequence 49; Application US/08634797
Patent No. 5851759
GENERAL INFORMATION:

APPLICANT: WEINER, AMY J.
TITLE OF INVENTION: HETEROIDPLEX TRACKING ASSAY (HTA) FOR
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Chiron Corporation
STREET: 4560 Horton Street - R440
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,797
FILING DATE: 19-APR-1996

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hardin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 1226.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-3274
TELEFAX: (510) 655-3542
TELEX: N/A
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-634-797-49

Query Match
Best Local Similarity 100.0%; Score 26; DB 2; Length 194;
Matches 26; Conservative 0; Pred. No. 0.00086;
Mismatched 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCATGCGCTTA 26
Db 1 GCAGAAAGCGCTAGCATGCGCTTA 26

RESULT 9
US-08-634-797-50

Sequence 50; Application US/08634797
Patent No. 5851759
GENERAL INFORMATION:

APPLICANT: WEINER, AMY J.
TITLE OF INVENTION: HETEROIDPLEX TRACKING ASSAY (HTA) FOR
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Chiron Corporation
STREET: 4560 Horton Street - R440
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916

NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Chiron Corporation
STREET: 4560 Horton Street - R440
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,797
FILING DATE: 19-APR-1996

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hardin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 1226.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-3274
TELEFAX: (510) 655-3542
TELEX: N/A
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-634-797-50

Query Match
Best Local Similarity 100.0%; Score 26; DB 2; Length 194;
Matches 26; Conservative 0; Pred. No. 0.00086;
Mismatched 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCATGCGCTTA 26
Db 1 GCAGAAAGCGCTAGCATGCGCTTA 26

RESULT 10
US-08-634-797-51
Sequence 51; Application US/08634797
Patent No. 5851759
GENERAL INFORMATION:

APPLICANT: WEINER, AMY J.
TITLE OF INVENTION: HETEROIDPLEX TRACKING ASSAY (HTA) FOR
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Chiron Corporation
STREET: 4560 Horton Street - R440
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,797
FILING DATE: 19-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hardin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 1226.001
TELECOMMUNICATION INFORMATION:

Query Match
Best Local Similarity 100.0%; Score 26; DB 2; Length 194;
Matches 26; Conservative 0; Pred. No. 0.00086;
Mismatched 0; Indels 0; Gaps 0;

TELEPHONE: (510) 601-3274
TELEFAX: (510) 655-3542
TELEX: N/A
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-634-797-51

Query Match 100.0%; Score 26; DB 2; Length 194;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCATGGCGTTA 26
|||||
DB 1 GCAGAAAGCGCTAGCATGGCGTTA 26

RESULT 11
US-08-634-797-52
Sequence 52, Application US/08634797
Patent No. 5851759
GENERAL INFORMATION:
APPLICANT: WEINER, AMY J.
TITLE OF INVENTION: HETERODUPLEX TRACKING ASSAY (HTA) FOR
TITLE OF INVENTION: GENOTYPING HCV
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street - RA40
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
TELEFAX: (510) 655-3542
TELEX: N/A
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-634-797-52

Query Match 100.0%; Score 26; DB 2; Length 194;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCATGGCGTTA 26
|||||
DB 1 GCAGAAAGCGCTAGCATGGCGTTA 26

RESULT 12

US-08-335-595-1
Sequence 1, Application US/08335595
Patent No. 5914228
GENERAL INFORMATION:
APPLICANT: VIERLING, JOHN M
APPLICANT: HU, KE-QIN
TITLE OF INVENTION: DIRECT DETECTION OF HEPATITIS C VIRUS
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: LYON & LYON
STREET: 611 WEST 6TH STREET
CITY: LOS ANGELES
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/335,595
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/175,473
FILING DATE:
APPLICATION NUMBER: US/07/758,862
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SCHNEIDER, CAROL A
REGISTRATION NUMBER: 34,923
REFERENCE/DOCKET NUMBER: 194/285
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-489-1600
TELEFAX: 213-955-0440
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 242 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-335-595-1

Query Match 100.0%; Score 26; DB 2; Length 242;
Best Local Similarity 100.0%; Pred. No. 0.00089;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCATGGCGTTA 26
|||||
DB 45 GCAGAAAGCGCTAGCATGGCGTTA 70

RESULT 13
US-09-034-205-26
Sequence 26, Application US/09034205
Patent No. 6194149
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance P.
APPLICANT: Neill, Bruce P.
TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING
STRUCTURE-BRIDGING OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30.
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,205
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Macknight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-03268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ. ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-034-205-26

Query Match
Best Local Similarity 100.0%; Score 26; DB 4; Length 244;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 GCAGAAAGCGCTAGCCATGGCGTTA 26
1 GCAGAAAGCGCTAGCCATGGCGTTA 26

RESULT 14
US-09-034-205-27
Sequence 27, Application US/09034205
Patent No. 6194149
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Nerl, Bruce P.
TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING
STRUCTURE-BRIDGING OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,205
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Macknight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-03268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ. ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-034-205-27

Query Match
Best Local Similarity 100.0%; Score 26; DB 4; Length 244;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 GCAGAAAGCGCTAGCCATGGCGTTA 26
1 GCAGAAAGCGCTAGCCATGGCGTTA 26

RESULT 15
US-09-034-205-28
Sequence 28, Application US/09034205
Patent No. 6194149
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Nerl, Bruce P.
TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING
STRUCTURE-BRIDGING OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,205
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Macknight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-03268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ. ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-034-205-28

Query Match
Best Local Similarity 100.0%; Score 26; DB 4; Length 244;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 GCAGAAAGCGCTAGCCATGGCGTTA 26
1 GCAGAAAGCGCTAGCCATGGCGTTA 26

Search completed: July 10, 2003, 20:27:39
Job time: 16.627 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:53:16 ; Search time 71.164 Seconds
(without alignments)
576.891 Million cell updates/sec

Title: US-10-087-631B-6
Perfect score: 26
Sequence: 1 gcagaaagcgtctagccatgacgtta 26

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCr_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/PCrTUS_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	26	100.0	26	9	US-10-087-631B-6
2	26	100.0	30	12	US-10-029-907-17
3	26	100.0	241	9	US-10-087-631B-10
4	26	100.0	241	9	US-10-087-631B-13
5	26	100.0	242	9	US-10-087-631B-15
6	26	100.0	244	10	US-09-825-574-26
7	26	100.0	244	10	US-09-825-574-27
8	26	100.0	244	10	US-09-825-574-28
9	26	100.0	281	9	US-09-940-925A-123
10	26	100.0	281	9	US-09-940-925A-125
11	26	100.0	281	9	US-09-940-925A-129
12	26	100.0	281	9	US-09-940-925A-129
13	26	100.0	281	9	US-09-941-193A-131
14	26	100.0	281	9	US-09-941-193A-123
15	26	100.0	281	9	US-09-941-193A-125
16	26	100.0	281	9	US-09-941-193A-128
17	26	100.0	281	9	US-09-941-193A-129
18	26	100.0	281	9	US-09-941-193A-131
19	26	100.0	282	9	US-09-940-925A-124

Query Match	Score	DB 9	Length 26	Best Local Similarity	100.0%	Pred. No. 0.0023	Mismatches	0	Indels	0	Gaps	0
1	GCAGAAAGCCTCTAGCCATGCGCTTA	26										
2	US-10-029-907-17											
3	Sequence 17, Appli											
4	Patent No. US20020142350A1											
5	GENERAL INFORMATION:											
6	APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.											
7	TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM											
8	FILE REFERENCE: 13/083											
9	CURRENT APPLICATION NUMBER: US/10/029,907											
10	CURRENT FILING DATE: 2001-12-21											

ALIGNMENTS

RESULT 1
US-10-087-631B-6
Sequence 6, Appli US/10087631B
Publication No. US20030034372A1

GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 26
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: sequence ST280 HCV-specifl
US-10-087-631B-6

Query Match 100.0%; Score 26; DB 9; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 GCAGAAAGCCTCTAGCCATGCGCTTA 26
|||||
US-10-029-907-17
Sequence 17, Appli
Patent No. US20020142350A1
GENERAL INFORMATION:
APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
FILE REFERENCE: 13/083
CURRENT APPLICATION NUMBER: US/10/029,907
CURRENT FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: 60/257,857
PRIOR FILING DATE: 2000-12-22
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 17
LENGTH: 30
TYPE: DNA
ORGANISM: HCV
US-10-029-907-17

Query Match 100.0%; Score 26; DB 12; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCCATGGCGTTA 26
DB 3 GCAGAAAGCGCTAGCCATGGCGTTA 28

RESULT 3

US-10-087-631b-10
Sequence 10, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:

APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999

CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01

NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 241
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: DNA sequence derived by
OTHER INFORMATION: amplification of HCV type 1 using primers ST280 and ST778
US-10-087-631b-10

Query Match 100.0%; Score 26; DB 9; Length 241;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCCATGGCGTTA 26
DB 1 GCAGAAAGCGCTAGCCATGGCGTTA 26

RESULT 4

US-10-087-631b-13
Sequence 13, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:

APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999

CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01

NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 13
LENGTH: 241
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Description of artificial sequence: Amplicon sequence derived fr
OTHER INFORMATION: QS HCV (HCV amplification control having binding sites for ST280,
OTHER INFORMATION: and ST253) using primers ST280 and ST778
US-10-087-631b-13

Query Match 100.0%; Score 26; DB 9; Length 241;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCCATGGCGTTA 26
DB 1 GCAGAAAGCGCTAGCCATGGCGTTA 26

RESULT 5

US-10-087-631b-15
Sequence 15, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:

APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999

CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01

NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 15
LENGTH: 242
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Description of artificial sequence: Amplicon derived from ICS
OTHER INFORMATION: HCV (HCV-specific amplification control) using ST280 and ST778
US-10-087-631b-15

Query Match 100.0%; Score 26; DB 9; Length 242;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCCATGGCGTTA 26
DB 1 GCAGAAAGCGCTAGCCATGGCGTTA 26

RESULT 6

US-09-825-574-26
Sequence 26, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:

APPLICANT: Lyamichev, Victor I.
Fors, Lance
Neri, Bruce P.

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FONS-02980
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-825-574-26

Query Match
Best Local Similarity 100.0%; Score 26; DB 10; Length 244;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTACGCGGCTTA 26
DB 1 GCAGAAAGCGCTACGCGGCTTA 26

RESULT 7
US-09-825-574-27
Sequence 27, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamlichev, Victor I.
Brow, Mary Ann D.
Fors, Lance P.
Neri, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FONS-02980
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"

SEQUENCE DESCRIPTION: SEQ ID NO: 27:
US-09-825-574-27

Query Match
Best Local Similarity 100.0%; Score 26; DB 10; Length 244;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTACGCGGCTTA 26
DB 1 GCAGAAAGCGCTACGCGGCTTA 26

RESULT 8
US-09-825-574-28
Sequence 28, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamlichev, Victor I.
Brow, Mary Ann D.
Fors, Lance P.
Neri, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FONS-02980
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-09-825-574-28

Query Match
Best Local Similarity 100.0%; Score 26; DB 10; Length 244;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTACGCGGCTTA 26
DB 1 GCAGAAAGCGCTACGCGGCTTA 26

RESULT 9
US-09-940-925A-123

Sequence 123, Application US/09940925A
Publication No. US20030054338A1
GENERAL INFORMATION:
APPLICANT: BROW, MARY ANN D.
OLIVE, DAVID M.
TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
PATHOGENS
NUMBER OF SEQUENCES: 165
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/940,925A
FILING DATE: 10-Jun-2002
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01756
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 123:
SEQUENCE CHARACTERISTICS:
LENGTH: 281 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 123:
US-09-940-925A-123

Query Match 100.0%; Score 26; DB 9; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 11 GCAGAAAGCGCTAGCCATGCGCTTA 26
11 GCAGAAAGCGCTAGCCATGCGCTTA 36

RESULT 10
US-09-940-925A-125
Sequence 125, Application US/09940925A
Publication No. US20030054338A1
GENERAL INFORMATION:
APPLICANT: BROW, MARY ANN D.
OLIVE, DAVID M.
TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
PATHOGENS
NUMBER OF SEQUENCES: 165
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/940,925A
FILING DATE: 10-Jun-2002
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01756
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 128:
SEQUENCE CHARACTERISTICS:
LENGTH: 281 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 128:
US-09-940-925A-128/c

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/940,925A
FILING DATE: 10-Jun-2002
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01756
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 125:
SEQUENCE CHARACTERISTICS:
LENGTH: 281 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 125:
US-09-940-925A-125

Query Match 100.0%; Score 26; DB 9; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 11 GCAGAAAGCGCTAGCCATGCGCTTA 26
11 GCAGAAAGCGCTAGCCATGCGCTTA 36

RESULT 11
US-09-940-925A-128/c
Sequence 128, Application US/09940925A
Publication No. US20030054338A1
GENERAL INFORMATION:
APPLICANT: BROW, MARY ANN D.
OLIVE, DAVID M.
TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
PATHOGENS
NUMBER OF SEQUENCES: 165
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/940,925A
FILING DATE: 10-Jun-2002
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01756
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 128:
SEQUENCE CHARACTERISTICS:
LENGTH: 281 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 128:
US-09-940-925A-128

US-09-940-925A-128

Query Match 100.0%; Score 26; DB 9; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
DB 271 GCAGAAAGCGTCTAGCCATGGCGTTA 246

RESULT 12

US-09-940-925A-129/C
; Sequence 129, Application US/09940925A
; Publication No. US20030054338A1

GENERAL INFORMATION:

APPLICANT: BROW, MARY ANN D.

LYAMICHEV, VICTOR I.

OLIVE, DAVID M.

TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
PATHOGENS

NUMBER OF SEQUENCES: 165

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MEDLEN & CARROLL

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/940.925A

FILING DATE: 10-Jun-2002

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: CARROLL, PETER G.

REGISTRATION NUMBER: 32,837

REFERENCE/DOCKET NUMBER: FORS-01756

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 129:

SEQUENCE CHARACTERISTICS:

LENGTH: 281 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 129:

US-09-940-925A-129

Query Match

Best Local Similarity 100.0%; Score 26; DB 9; Length 281;
Pred. No. 0.0027;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
DB 271 GCAGAAAGCGTCTAGCCATGGCGTTA 246

RESULT 13

US-09-940-925A-131/C
; Sequence 131, Application US/09940925A
; Publication No. US20030054338A1

GENERAL INFORMATION:

APPLICANT: BROW, MARY ANN D.

LYAMICHEV, VICTOR I.

OLIVE, DAVID M.

TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF

PATHOGENS

NUMBER OF SEQUENCES: 165

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MEDLEN & CARROLL

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/940.925A

FILING DATE: 10-Jun-2002

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: CARROLL, PETER G.

REGISTRATION NUMBER: 32,837

REFERENCE/DOCKET NUMBER: FORS-01756

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 131:

SEQUENCE CHARACTERISTICS:

LENGTH: 281 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 131:

US-09-940-925A-131

Query Match

Best Local Similarity 100.0%; Score 26; DB 9; Length 281;
Pred. No. 0.0027;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
DB 271 GCAGAAAGCGTCTAGCCATGGCGTTA 246

RESULT 14

US-09-941-193A-123
; Sequence 123, Application US/09941193A
; Publication No. US20030108873A1

GENERAL INFORMATION:

APPLICANT: BROW, MARY ANN D.

LYAMICHEV, VICTOR I.

OLIVE, DAVID M.

TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
PATHOGENS

NUMBER OF SEQUENCES: 165

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MEDLEN & CARROLL

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/941.193A

FILING DATE: 28-Aug-2001

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: CARROLL, PETER G.

Db 11 GCAGAAAGCGCTAGCCATGCGCTTA 36
 Search completed: July 11, 2003, 15:01:56
 Job time: 71.164 secs

REGISTRATION NUMBER: 32,837
 REFERENCE/DOCKET NUMBER: F0RS-01756
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 123:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 281 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 123:
 US-09-941-193A-123

Query Match 100.0%; Score 26; DB 9; Length 281;
 Best Local Similarity 100.0%; Pred. No. 0.0027;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCGCTAGCCATGCGCTTA 26
 Db 11 GCAGAAAGCGCTAGCCATGCGCTTA 36

RESULT 15
 US-09-941-193A-125

Sequence 125: Application US/09941193A
 Publication No. US20030108873A1
 GENERAL INFORMATION:

APPLICANT: BROW, MARY ANN D.
 LYAMICHEV, VICTOR I.
 OLIVE, DAVID M.
 TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
 PATHOGENS
 NUMBER OF SEQUENCES: 165
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MEDLEN & CARROLL
 STREET: 220 MONTGOMERY STREET, SUITE 2200
 CITY: SAN FRANCISCO
 STATE: CALIFORNIA
 COUNTRY: UNITED STATES OF AMERICA
 ZIP: 94104

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/941,193A
 FILING DATE: 28-Aug-2001
 CLASSIFICATION: <unknown>

ATTORNEY/AGENT INFORMATION:
 NAME: CARROLL, PETER G.
 REGISTRATION NUMBER: 32,837
 REFERENCE/DOCKET NUMBER: F0RS-01756
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 125:

SEQUENCE CHARACTERISTICS:
 LENGTH: 281 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 125:
 US-09-941-193A-125

Query Match 100.0%; Score 26; DB 9; Length 281;
 Best Local Similarity 100.0%; Pred. No. 0.0027;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCGCTAGCCATGCGCTTA 26

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:13:45 ; Search time 467.474 Seconds
(without alignments)
900.761 Million cell updates/sec

Title: US-10-087-631b-6

Perfect score: 26

Sequence: 1 gcagaagcgtctacgcatgcycta 26

Scoring table: IDENTITY_NUC

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20.2	77.7	549	12	BG443799 GA_Ea002
2	20.2	77.7	603	10	BE055455 GA_Ea000
3	20.2	77.7	610	13	BM358998 GA_Ea001
4	20.2	77.7	634	14	BQ404294 GA_Ed006
5	20.2	77.7	638	14	BQ401354 GA_Ed000
6	20.2	77.7	642	12	BG441998 GA_Ea001

7	20.2	77.7	694	12	BG440823 GA_Ea001
8	20.2	77.7	779	10	BE055455 GA_Ea000
9	20.2	77.7	840	10	BE278075 GA_Eb003
10	18.6	71.5	609	14	BQ839191 WHEA163.D
11	18.6	71.5	688	14	BQ405841 GA_Ed008
12	18.4	70.8	313	10	AM896937 RC4-NN005
13	18.2	70.0	240	9	AJ461009 AJ461009
14	18.2	70.0	752	17	A2559641 RPCI-23-2
15	18	69.2	377	12	BF911205 IL2-UTR007
16	18	69.2	383	12	BF911367 IL2-UTR007
17	18	69.2	396	12	BF911282 IL2-UTR007
18	18	69.2	402	12	BF911293 IL2-UTR007
19	18	69.2	415	12	BF911279 IL2-UTR007
20	18	69.2	450	10	BE475077 sp72a05.Y
21	18	69.2	564	13	B1470037 sap7a6h05.
22	18	69.2	590	14	BQ742735 sag55d09.
23	18	69.2	668	12	BG398045 AGENCOURT
24	18	69.2	883	14	BQ686524 AGENCOURT
25	17.8	68.5	326	14	BQ507679 EST615094
26	17.8	68.5	393	9	AU068851 AU068851
27	17.8	68.5	403	9	AU076005 AU076005
28	17.8	68.5	421	9	AU162743 AU162743
29	17.8	68.5	435	14	D49161 RICS15861A
30	17.8	68.5	595	17	A2865010 2M0174K22
31	17.6	67.7	344	9	A1255537 u155g04.Y
32	17.6	67.7	520	13	B1188143 B1188143
33	17.6	67.7	520	13	BM168779 EST571302
34	17.6	67.7	549	10	BE632435 uv56c06.Y
35	17.6	67.7	563	12	BG146531 mad93b06.
36	17.6	67.7	564	13	B1184088 B1184088
37	17.6	67.7	566	13	BM660882 952042F08
38	17.6	67.7	637	10	BB625273 BB625273
39	17.6	67.7	662	17	A2612987 1M0441P12
40	17.6	67.7	695	17	BH558329 BOGHN27TR
41	17.6	67.7	726	14	BM937432 UT-M-CD1
42	17.6	67.7	735	17	BH520932 BOGGA53TF
43	17.6	67.7	807	17	BH597113 BOGDY77TR
44	17.6	67.7	812	13	B1218614 602937835
45	17.6	67.7	833	17	BH490369 BOHHE31TF

ALIGNMENTS

RESULT 1
LOCUS BG443799 549 bp mRNA linear EST 15-MAR-2001
DEFINITION GA_Ea0022C04f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_Ea0022C04f, mRNA sequence.

ACCESSION BG443799.1 GI:13353451

VERSION
KEYWORDS
SOURCE
ORGANISM

Gossypium arboreum.

Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.

1 (bases 1 to 549)

Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
, D., Wood, T.C., Leslie, A. and Wilkins, T.A.

An integrated analysis of the genetics, development, and evolution
of the cotton fiber

Unpublished (2000)

Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: twing@clemson.edu
Seq primer: TAAATGACATCACTATAGCG
High quality sequence stop: 542.

FEATURES

Location/Qualifiers

JOURNAL
COMMENT

of the cotton fiber
unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total High Quality bases = 593
Seq primer: TAAATGAGCTCACTATAGGG
High quality sequence stop: 632.
High quality sequence stop: 632.
Location/Qualifiers
1. 634
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 188 a 124 c 140 g 176 t 6 others
ORIGIN

Query Match 77.7%; Score 20.2; DB 14; Length 634;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGCGCTT 25
|||||
DB 28 GCAGAAAGCGTCTAGCCATGACTTT 52

RESULT 5
LOCUS BQ401354 638 bp mRNA linear EST 22-MAR-2002
DEFINITION GA_Ed0001A07f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboeum cDNA clone GA_Ed0001A07f, mRNA sequence.
ACCESSION BQ401354
VERSION BQ401354.1 GI:21089041
KEYWORDS EST
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 638)
Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
D., Wood, T.C., Leslie, A. and Wilkins, T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total High Quality bases = 584
Seq primer: TAAATGAGCTCACTATAGGG
High quality sequence stop: 635.
Location/Qualifiers
1. 638
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
SOURCE

/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 188 a 125 c 144 g 181 t
ORIGIN

Query Match 77.7%; Score 20.2; DB 14; Length 638;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGCGCTT 25
|||||
DB 22 GCAGAAAGCGTCTAGCCATGACTTT 46

RESULT 6
LOCUS BG441998 642 bp mRNA linear EST 15-MAR-2001
DEFINITION GA_Ea0015G07f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboeum cDNA clone GA_Ea0015G07f, mRNA sequence.
ACCESSION BG441998
VERSION BG441998.1 GI:13351650
KEYWORDS EST
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 642)
Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
D., Wood, T.C., Leslie, A. and Wilkins, T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAAATGAGCTCACTATAGGG
High quality sequence stop: 642.
Location/Qualifiers
1. 642
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

FEATURES
SOURCE

BASE COUNT 187 a 128 c 146 g 181 t
ORIGIN

Query Match 77.7%; Score 20.2; DB 12; Length 642;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGCGCTT 25
|||||
DB 31 GCAGAAAGCGTCTAGCCATGACTTT 55

RESULT 7
LOCUS BG440823 694 bp mRNA linear EST 15-MAR-2001
DEFINITION GA_Ea0010F07f Gossypium arboreum 7-10 dpa fiber library Gossypium

ACCESSION	BG440823		airdium cdna clone GA_Ea0010F07f, mRNA sequence.
VERSION	BG440823.1	GI:13350475	
KEYWORDS	EST.		
SOURCE	Gossypium arboreum.		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		
REFERENCE	1 (bases 1 to 694)		
AUTHORS	Wing,R.A., Fritsch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry D., Wood,T.C., Leslie,A. and Wilkins,T.A.		
TITLE	An integrated analysis of the genetics, development, and evolution of the cotton fiber		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Wing RA Clemson University Genomics Institute 100 Jordan Hall, Clemson, SC 29634, USA Tel.: 864 656 7288 Fax: 864 656 4293 Email: rwing@clemson.edu Seq primer: TAATACACTCACTAAGGG High quality sequence stop: 353. Location/Qualifiers		
FEATURES	1..694		
source	/organism="Gossypium arboreum"		
	/strain="AKA"		
	/cultivar="8400"		
	/db_xref="taxon:29729"		
	/clone="GA_Ea0010F07f"		
	/clone_1lb="Gossypium arboreum 7-10-dpa fiber library"		
	/tissue_type="fibers isolated from bolls harvested 7-10 dpa"		
	/lab_host="E. coli"		
	/note=Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	191 a	146 c	154 g
ORIGIN	203 t		
Query Match	77.7%; Score 20.2; DB 12;		
Best Local Similarity	88.0%; Pred. No. 1.1e+02;		
Matches	22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		
Oy	1 GCAGAAGCGTCTAGCCATGGCGTT 25		
Db	18 GCAGAAGCCTCTAGCATGACTTT 42		
RESULT 8			
BE055545		779 bp	mRNA linear EST 07-MAR-2001
LOCUS	GA_Ea0004L16f Gossypium arboreum 7-10 dpa fiber library Gossypium		
DEFINITION	arborum cdna clone GA_Ea0004L16f, mRNA sequence.		
ACCESSION	BE055545		
VERSION	BE055545		
KEYWORDS	EST.		
SOURCE	Gossypium arboreum.		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		
REFERENCE	1 (bases 1 to 779)		
AUTHORS	Wing,R.A., Fritsch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry D., Wood,T.C., Leslie,A. and Wilkins,T.A.		
TITLE	An integrated analysis of the genetics, development, and evolution of the cotton fiber		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Wing RA Clemson University Genomics Institute 100 Jordan Hall, Clemson, SC 29634, USA Tel.: 864 656 7288		

FEATURES	source	1. 779	/organism="Gossypium arboreum"		
			/strain="AKA"		
			/cultivar="8400"		
			/db_xref="taxon:29729"		
			/clone="GA_Ea0004L16f"		
			/clone_lib="Gossypium arboreum 7-10 dpa fiber library"		
			/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"		
			/lab_host="E. coli"		
			/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	220 a	158 c	177 g	217 t	7 others
ORIGIN					
Query Match	Similarity	88.0%	Pred. No. 1.1e+02;		
Best Local	Matches	22;	Conservative	0;	Mismatches 3; Indels 0; Gaps 0;
QY	1	GCAGAAAGCGTCTAGCCATGCGCCTT	25		
Db	27	GCAGAAAGCGTCTAGCCATGACTTT	51		
RESULT 9	BF278075	840 bp	mRNA	linear	EST 07-MAR-2001
LOCUS	GA_EB0038A09f	Gossypium arboreum 7-10 dpa fiber library	Gossypium		
DEFINITION	arborescens cDNA clone GA_EB0038A09f, mRNA sequence.				
ACCESSION	BF278075				
VERSION	BF278075.1	GI:11209061			
KEYWORDS	EST.				
SOURCE	Gossypium arboreum.				
ORGANISM	Gossypium arboreum				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
	Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.				
	1. (bases 1 to 840)				
REFERENCE	Wing,R.A., Fritsch,D., Yu,Y., Maiti,D., Rambo,T., Simmons,J., Henry				
AUTHORS	,D., Wood,T.C., Leslie,A. and Wilkins,T.A.				
TITLE	An integrated analysis of the genetics, development, and evolution				
	of the cotton fiber				
	Unpublished (2000)				
JOURNAL	Contact: Wing RA				
COMMENT	Clemson University Genomics Institute				
	Clemson University				
	100 Jordan Hall, Clemson, SC 29634, USA				
	Tel: 864 656 7288				
	Fax: 864 656 4293				
	Email: rwing@clemson.edu				
	Seq primer: TAATACGACTCATTATGAG				
	High quality sequence start: 4				
	High quality sequence stop: 535.				
FEATURES	Location/Qualifiers				
source	1. 840				
	/organism="Gossypium arboreum"				
	/strain="AKA"				
	/cultivar="8400"				
	/db_xref="taxon:29729"				
	/clone="GA_EB0038A09f"				
	/clone_lib="Gossypium arboreum 7-10 dpa fiber library"				
	/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"				
	/lab_host="E. coli"				
	/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"				
BASE COUNT	198 a	194 c	148 g	298 t	2 others
ORIGIN					
Query Match	77.7%	Score 20.2;	DB 12;	Length 840;	

Best Local Similarity 88.0%; Pred. No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTT 25
DB 30 GCAGAAAGCGTCTAGCCATGGCGTT 54

RESULT 10
BO839191

LOCUS BO839191 609 bp mRNA linear EST 08-AUG-2002
DEFINITION wheat CS whole plant cDNA library Triticum aestivum cDNA clone WHE4163_D05_G09, mRNA sequence.

ACCESSION

VERSION BO839191.1 GI:22143513

KEYWORDS

SOURCE bread wheat.

ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae; 1 (bases 1 to 609)

AUTHORS Anderson O.D., Akhunov, E., Chao, S., Crossman, C., Deal, K., Dvorak, J., Lazo, G.R., Rausch, C.J., Wilson, C., and Woo, J.

TITLE The structure and function of the expressed portion of the wheat genomes - Chinese Spring whole plant cDNA library

JOURNAL Unpublished (2002)

COMMENT Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818

EMAIL: andersn@pw.usda.gov

Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: SK primer.

FEATURES

SOURCE

location/Qualifiers

1..609

/organism="Triticum aestivum"

/cultivar="Chinese Spring"

/db_xref="taxon:4565"

/clone="WHE4163_D05_G09"

/clone_lib="wheat CS whole plant cDNA library"

/tissue_type="Roots, leaves, crown, stem and sheath"

/dev_stage="Adult"

/lab_host="E. coli SOLR"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid

bluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Plant

tissues from wheat cv. CS grown to full tillering stage in

greenhouse were collected at University of California,

Davis (Jan Dvorak's lab, K. Deal and E. Akhunov). Total

RNA was prepared from leaves (young leaf and third leaf),

whole roots, crown, stem and sheath tissues, and then

equal quantities of RNA were pooled from the these

samples. PolyA was purified from the pooled RNA, a cDNA

library was made, and the cDNA clones were in vivo excised

to give bluescript SK(-) phagemids in J. Dvorak's lab (E.

Akhunov, J. Dvorak) at the University of California,

Davis. Colony plating, plasmid DNA preparations and DNA

sequencing were performed in the OD Anderson lab (all

other authors)."

BASE COUNT 109 a 183 c 231 g 86 t

ORIGIN

Query Match

Best Local Similarity 84.0%; Pred. No. 5.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTT 25

DB 435 GCAGAAAGCGTCTAGCCATGGCGTT 459

RESULT 11
BQ405841

LOCUS BQ405841 688 bp mRNA linear EST 22-MAY-2002
DEFINITION GA_Ed0087E10f Gossypium arboreum 7-10 dpa fiber library Gossypium arboreum cDNA clone GA_Ed0087E10f, mRNA sequence.

ACCESSION

VERSION BQ405841

KEYWORDS

SOURCE EST.

ORGANISM Gossypium arboreum.

Gossypium arboreum. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.

REFERENCE 1 (bases 1 to 688)

AUTHORS Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry

D., Wood, T.C., Leslie, A. and Wilkins, T.A.

An integrated analysis of the genetics, development, and evolution of the cotton fiber

Unpublished (2000)

JOURNAL

COMMENT Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

EMAIL: rwing@clemson.edu

Total High Quality bases = 442

Seq primer: TATACGACTCCTATGAGG

High quality sequence start: 33

High quality sequence stop: 681.

FEATURES

SOURCE

location/Qualifiers

1..688

/organism="Gossypium arboreum"

/strain="AKA"

/cultivar="8400"

/db_xref="taxon:29729"

/clone="GA_Ed0087E10f"

/clone_lib="Gossypium arboreum 7-10 dpa fiber library"

/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"

/lab_host="E. coli"

/note="Vector: pRK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 201 a 134 c 149 g 204 t

ORIGIN

Query Match

Best Local Similarity 84.0%; Pred. No. 5.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCCATGGCGTT 25

DB 12 GCAGAAAGCGTCTAGCCATGGCGTT 36

RESULT 12
AM896937

LOCUS AM896937 313 bp mRNA linear EST 24-MAY-2000

DEFINITION RC4-NN0055-060400-011-E09 NN0055 Homo sapiens cDNA, mRNA sequence.

ACCESSION AM896937

VERSION AM896937.1 GI:8061142

KEYWORDS

SOURCE EST.

ORGANISM human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,

Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.,
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rita Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=6t2-R64-NN0055-060
400-011-109&t3=2000-04-06&t4=1)
Seq primer: puc 18 forward
High quality sequence, start: 21
High quality sequence, stop: 313.
Location/Qualifiers
1. 313

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN0055"
/dev_stage="Adult"

BASE COUNT

64 a 85 c 83 g 81 t

ORIGIN

Query Match 70.0%; Score 18.4; DB 10; Length 313;
Best Local Similarity 95.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 7 AGCGCTAGCCATGCGCTTA 26
db 68 AGCGCTAGCCATGCGCTGA 87

RESULT 13
AJ461009 240 bp mRNA linear EST 24-MAY-2002
LOCUS AJ461009 S00002 Hordeum vulgare cDNA S0000200049A08F1, mRNA
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Hordeum vulgare.
EST.
Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
1 (bases 1 to 240)
Saren, A.-M., Tanskanen, J., Paulin, L. and Schulman, A. H.
Barley EST 5
Unpublished (2002)
Contact: Schulman AH
Institute of Biotechnology
University of Helsinki
P.O. Box 56 (Vilkkinkaari 6A); University of Helsinki FIN-00014,
Finland.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source
1. 240
Location/Qualifiers
/organism="Hordeum vulgare"
/cultivar="Saana"
/db_xref="taxon:4513"
/clone="S0000200049A08F1"
/clone_lib="S00002"

/dev stage="Embryo"
/note="1 day after pollination"
BASE COUNT 63 a 59 c 57 g 61 t
ORIGIN

Query Match 70.0%; Score 18.2; DB 9; Length 240;
Best Local Similarity 87.0%; Pred. No. 4.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCCATGCGC 23
db 45 GAAGAAAGCGCTAGCCACGCGC 67

RESULT 14
AJ559641 752 bp DNA linear GSS 20-NOV-2000
LOCUS AJ559641
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 752)
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akiret
J., Levins, M., McGann, S., Ysegaire, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other GSSs: RPCI-23-205015.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 205 row: 0 column: 15
Seq primer: T7
Class: BAC ends.

FEATURES
source
1. 752
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-205015"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1;
EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 216 a 175 c 159 g 202 t
ORIGIN

Query Match 70.0%; Score 18.2; DB 17; Length 752;
Best Local Similarity 87.0%; Pred. No. 9.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCCATGCGC 23

Db 701 GCAGAAAGTGTCTAGTCAGGCG 723

RESULT 15

BF912055/c 377 bp mRNA linear EST 18-JAN-2001
 LOCUS IL2-UT0073-121100-232-D04 UT0073 Homo sapiens cDNA, mRNA sequence.
 DEFINITION
 ACCESSION BF912055
 VERSION BF912055.1 GI:12303513
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 377)

REFERENCE
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

FEATURES
 SOURCE This sequence was derived from the FAPESP/LICR Human Cancer Genome
 project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL2&ct=IL2-UT0073-
 121100-232-D04&ct3=2000-11-12&ct4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 349.
 Location/Qualifiers

1. 377
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="UT0073"
 /dev_stage="Adult"
 /note="Organ: uterus,tumor; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 98 a 66 c 76 g 136 t 1 others
 ORIGIN

Query Match 69.2%; Score 18; DB 12; Length 377;
 Best Local Similarity 80.8%; Pred.No. 7.7e+02;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCTGTAGCCATGCGCTTA 26
 ||||| ||||| ||||| |||||
 Db 322 GCAGAAAGCTACACCCATGTAGTTA 297

Search completed: July 11, 2003, 02:25:50
 Job time : 474.474 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 59.0831 Seconds

(without alignments)
1067.241 Million cell updates/sec

Title: US-10-087-631B-7

Perfect score: 28
Sequence: 1 gcaagcaccctatcgagcgatcacca 28

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries.

Database :

N.Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	28	18	AAAT67194
2	28	100.0	28	19	AAV59059
3	28	100.0	28	22	AAH25414
4	28	100.0	57	14	AAAG3223
5	28	100.0	186	17	AAAT11272
6	28	100.0	209	24	ABLA1919
7	28	100.0	232	19	AAV70460
8	28	100.0	232	24	ABLA6070
9	28	100.0	239	19	AAV70455

C 10	28	100.0	239	19	AAV70459	Partial sequence o
C 11	28	100.0	239	24	ABLA6065	Hepatitis C virus
C 12	28	100.0	239	24	ABLA6069	Hepatitis C virus
C 13	28	100.0	240	19	AAV70457	Partial sequence o
C 14	28	100.0	240	19	AAV70456	Partial sequence o
C 15	28	100.0	240	19	AAV70458	Partial sequence o
C 16	28	100.0	240	24	ABLA6066	Hepatitis C virus
C 17	28	100.0	240	24	ABLA6067	Hepatitis C virus
C 18	28	100.0	244	24	ABLA6068	Hepatitis C virus
C 19	28	100.0	244	19	AAV70454	Partial sequence o
C 20	28	100.0	244	19	AAV70449	HCV subtype Ia PCR
C 21	28	100.0	244	19	AAV70450	HCV subtype Ib PCR
C 22	28	100.0	244	19	AAV70451	HCV subtype 2c PCR
C 23	28	100.0	244	19	AAV70452	HCV subtype 3a PCR
C 24	28	100.0	244	24	ABLA6059	Hepatitis C virus
C 25	28	100.0	244	24	ABLA6060	Hepatitis C virus
C 26	28	100.0	244	24	ABLA6061	Hepatitis C virus
C 27	28	100.0	244	24	ABLA6062	Hepatitis C virus
C 28	28	100.0	244	24	ABLA6064	Hepatitis C virus
C 29	28	100.0	252	13	AAO31071	HCV-1 genotype GII
C 30	28	100.0	252	13	AAO31072	HCV-1 genotype GII
C 31	28	100.0	252	13	AAO31078	HCV-1 genotype GII
C 32	28	100.0	252	13	AAO31079	HCV-1 genotype GII
C 33	28	100.0	252	13	AAO31067	HCV-1 genotype GI
C 34	28	100.0	252	13	AAO31067	HCV-1 genotype GI
C 35	28	100.0	252	13	AAO31068	HCV-1 genotype GI
C 36	28	100.0	252	13	AAO31069	HCV-1 genotype GI
C 37	28	100.0	252	13	AAO31070	HCV-1 genotype GI
C 38	28	100.0	252	13	AAO31080	HCV-1 genotype GIV
C 39	28	100.0	252	13	AAO31081	HCV-1 genotype GIV
C 40	28	100.0	256	13	AAO32981	HCV EI 5' non-codi
C 41	28	100.0	256	20	AAH84004	HCV EI 5' non-codi
C 42	28	100.0	256	20	AAH16761	Hepatitis C virus
C 43	28	100.0	267	24	ABN79975	Hepatitis C virus
C 44	28	100.0	278	24	ABN79973	Hepatitis C virus
C 45	28	100.0	279	24	ABA01119	Hepatitis C virus

ALIGNMENTS

RESULT 1	
ID	AAAT67194
AAAT67194	standard; DNA; 28 BP.
AC	AAAT67194;
AC	13-FEB-1998 (first entry)
DT	
XX	Hepatitis C virus (HCV) RNA amplification primer ST778AA.
XX	Hepatitis C virus; HCV; ST778AA; reverse transcription PCR; RT-PCR;
XX	detection; PCR primer; ss.
KW	
KW	Synthetic.
OS	
XX	EP776981-A2.
PN	04-JUN-1997.
XX	
PD	
XX	21-NOV-1996; 96EP-0118704.
PF	
XX	29-NOV-1995; 95US-0007739.
PR	
XX	(HOFF) HOFFMANN LA ROCHE & CO AG F.
PA	
XX	Tsang SY;
PI	
XX	WPI; 1997-291296/27.
DR	
XX	Oligonucleotide primers for hepatitis C virus RNA amplification
PT	by polymerase chain reaction
XX	

PS Claim 1, Page 12, 16pp; English.

CC This downstream primer ST778AA is used in the amplification of the

CC Hepatitis C virus (HCV) RNA by reverse transcription PCR. This is used

CC to amplify a 250 base pair product from the 5' untranslated region of

CC the HCV genome. This can be used to detect HCV in a sample with increased

CC sensitivity. Amplification of HCV nucleic acid using this primer is up to

CC 100 times more efficient than amplification with prior art primers.

XX

SQ Sequence 28 BP; 10 A; 10 C; 5 G; 3 T; 0 other;

Query Match 100.0%; Score 28; DB 18; Length 28;

Best Local Similarity 100.0%; Pred. No. 0.0028;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCTATCAGCGATACCA 28

DB 1 GCAAGCACCTATCAGCGATACCA 28

RESULT 2

AAV59059 standard; DNA; 28 BP.

AC AAV59059;

DT 07-JAN-1999 (first entry)

DE Primer ST778AA for HCV fragment.

XX PCR primer; HCV; nucleic acid amplification; ss.

OS Synthetic.

OS Human cytomegalovirus.

XX

Key Location/Qualifiers

FT modified_base 25 /tag= a

FT /note= "optionally benzylated"

FT modified_base 27 /tag= b

FT /note= "optionally benzylated"

FT modified_base 28 /tag= c

FT /note= "optionally benzylated; methylated, or nitrobenzylated"

PN EP866071-A2.

XX

PD 23-SEP-1998.

XX

PF 12-MAR-1998; 98EP-0104461

XX

PR 20-MAR-1997; 97US-00411271

XX

PA (HOFF) HOFFMANN LA ROCHE & CO AG F.

XX

PI Will SG, Young KKY;

XX

DR WPI; 1998-482929/42.

XX

PT Oligo-nucleotide(s) containing N-substituted nucleotide useful as

XX primers for nucleic acid amplification

XX

PS Example 6; Page 16; 38pp; English.

XX

CC This sequence represents a primer for a fragment of HCV, and is an

CC example of an oligonucleotide of the invention. The oligonucleotides of

CC the invention are of the formula 5'-SI-Nu-3' or 5'-SI-Nu-S2-3', where

CC SI is a sequence of 5-50 nucleotides; S2 is a sequence of 1-3

CC nucleotides; and Nu is a nucleotide with a purine or pyrimidine base

CC having an exocyclic amino group substituted by CHR1R2; R1, R2 are H,

CC 1-10C alkyl, alkoxy, optionally substituted phenyl, phenoxy or optionally

CC substituted naphthyl. The oligonucleotides are useful as primers for

CC nucleic acid amplification, preferably by polymerase chain reaction. Use

CC of the modified primers reduces non-specific amplification, especially

CC primer dimer formation, with a concomitant increase in the yield of the

CC intended target.

XX

SQ Sequence 28 BP; 10 A; 10 C; 5 G; 3 T; 0 other;

Query Match 100.0%; Score 28; DB 19; Length 28;

Best Local Similarity 100.0%; Pred. No. 0.0028;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCTATCAGCGATACCA 28

DB 1 GCAAGCACCTATCAGCGATACCA 28

RESULT 3

AAH25414 standard; DNA; 28 BP.

AC AAH25414;

DT 22-AUG-2001 (first entry)

DE Reverse PCR primer used to amplify a HCV DNA fragment.

XX Magnetic glass particle; nucleic acid purification; PCR primer; ss.

OS Hepatitis C virus.

XX

Key Location/Qualifiers

FT modified_base 28 /tag= a

FT /note= "derivatisation with a p-(t-butyl)benzyl residue"

PN WO200137291-A1.

XX

PD 25-MAY-2001.

XX

PF 17-NOV-2000; 2000MO-EP11459.

XX

PR 17-NOV-1999; 99EP-0122853.

XX

PR 12-MAY-2000; 2000EP-0110165.

XX

PA (HOFF) ROCHE DIAGNOSTICS GMBH.

XX

PI Weindel K, Riedling M, Geiger A;

XX

DR WPI; 2001-381247/40.

XX

PT Novel composition of magnetic glass particles for purification of DNA

XX or RNA in automated processes

XX

PS Example 7; Page 98; 105pp; English.

XX

CC The specification describes a composition of magnetic glass particles,

CC which contain at least one magnetic object with a mean diameter between

CC 5-500 nm. The composition is useful for the purification of nucleic

CC acids. The composition can be used to process large quantities of

CC nucleic acid samples, because it does not involve the particles being

CC centrifuged or the fluids being drawn through glass fiber filters.

CC PCR primers AAH25413-14 were used to amplify HCV DNA fragments. The

CC amplified fragment can be purified using the method of the invention.

XX

SQ Sequence 28 BP; 10 A; 10 C; 5 G; 3 T; 0 other;

Query Match 100.0%; Score 28; DB 22; Length 28;

Best Local Similarity 100.0%; Pred. No. 0.0028;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCTATCAGCGATACCA 28

DB 1 GCAAGCACCTATCAGCGATACCA 28

DB 1 GCAAGCACCCCTATCAGCGAGTACCAACA 28

RESULT 4
AA063223/C
ID AA063223 standard; RNA; 57 BP.
XX
AC AA063223;
XX
DT 13-JUN-1994 (first entry)
XX
DE Hepatitis C virus probe target region.
XX
KW Detection; HCV; 11:2 probe design.
XX
OS Hepatitis C virus.
XX
PN W09324656-A.
XX
PD 09-DEC-1993.
XX
PE 24-MAY-1993; 93WO-US04863.
XX
PR 29-MAY-1992; 92US-0891543.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Carrino JJ, Marshall RL, Sustachek JC;
XX
DR WPI; 1993-405844/50.
XX
PT Amplifying known RNA target for use in diagnosis of HIV and HCV.
PT Infection - by treating sample RNA with oligo-nucleotide probe,
PT extending probe by reverse transcription of target, dissociating
PT probe from target, hybridising 2nd probe with 1st, etc.
XX
PS Example 8; Page 26; 49pp; English.
XX
CC The sequence is that of the target region of probes (AA053257-053260)
CC used in the detection of hepatitis C virus (HCV) using a 11:2 probe
CC design. It corresponds to positions 246-302 of the 5' UTR of the
CC HcHVR sequence.
XX
SQ Sequence 57 BP; 9 A; 9 C; 23 G; 16 T; 0 other;

Query Match 100.0%; Score 28; DB 14; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.0031;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAAGCACCCCTATCAGCGAGTACCAACA 28
DB 54 GCAAGCACCCCTATCAGCGAGTACCAACA 27

RESULT 5
AAT11272
ID AAT11272 standard; RNA; 186 BP.
XX
AC AAT11272;
XX
DT 26-JUN-1996 (first entry)
XX
DE Hepatitis C virus partial 5'-UTR antisense RNA AS15.
XX
KW Antisense; therapy; complementary; HCV; 5'-untranslated region;
KW hepatitis C virus; inhibition; infection; treatment; stem-loop;
KW clone 2-1; ss.
XX
OS Hepatitis C virus.
XX
PN JP07303485-A.
XX
PD 21-NOV-1995.

XX 13-MAY-1994; 94JP-0124609.
PF
XX
PR 13-MAY-1994; 94JP-0124609.
XX
PA (TOFU) TONEN CORP.
XX
DR WPI; 1996-035187/04.
XX
PT Hepatitis C virus (HCV) anti-sense RNA - inhibits HCV structural
PT gene expression in vivo for treatment of HCV infection
XX
PS Claim 2; Page 10; 12pp; Japanese.
XX
CC The present sequence is a specifically claimed example of RNA that
CC is complementary (i.e. antisense) to part of the 5'-untranslated
CC region of the hepatitis C virus genome sequence contained in clone
CC 2-1. The 5'-UTR includes several stem-loop sequences. The antisense
CC RNA is useful for inhibiting expression of HCV structural genes and
CC thereby inhibiting viral replication in vivo. The antisense therapy
CC can be used in addition to conventional interferon treatment of HCV
CC infections.
XX
SQ Sequence 186 BP; 41 A; 65 C; 48 G; 32 U; 0 other;

Query Match 100.0%; Score 28; DB 17; Length 186;
Best Local Similarity 89.3%; Pred. No. 0.0037;
Matches 25; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAAGCACCCCTATCAGCGAGTACCAACA 28
DB 10 GCAAGCACCCCTATCAGCGAGTACCAACA 37

RESULT 6
ABL41919/C
ID ABL41919 standard; RNA; 209 BP.
XX
AC ABL41919;
XX
DT 11-JUN-2002 (first entry)
XX
DE DNA comprising consensus endodogmatic sequence/exodogmatic sequence.
XX
KW cellular organism; pathogen; retroviral particle; probe; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT 5..49
FT stem_loop /*tag= a
FT misc_structure 5..7
FT /*tag= b
FT /*note= "these bases bind to bases 47-49"
FT 10..14
FT /*tag= c
FT /*note= "these bases bind to bases 39-43"
FT 39..43
FT /*tag= d
FT /*note= "these bases bind to bases 10-14"
FT 92..104
FT /*tag= e
FT misc_structure 92..96
FT /*tag= f
FT /*note= "these bases bind to bases 100-104"
FT 100..104
FT misc_structure /*tag= g
FT /*note= "theses bases bind to bases 92-96"
FT 105..118
FT stem_loop /*tag= h
FT misc_structure 105..108
FT /*tag= i
FT /*note= "theses bases bind to bases 115-118"

```

FT misc-structure 116..118
FT /tag- j
FT /note= "theses bases bind to bases 105-108"
FT stem_loop 153..208
FT /tag- k
FT misc-structure 153..157
FT /tag- l
FT /note= "theses bases bind to bases 204-208"
FT misc-structure 158..160
FT /tag- m
FT /note= "theses bases bind to bases 193-195"
FT misc-structure 162..167
FT /tag- n
FT /note= "theses bases bind to bases 186-191"
FT misc-structure 171..173
FT /tag- o
FT /note= "theses bases bind to bases 182-184"
FT misc-structure 182..184
FT /tag- p
FT /note= "theses bases bind to bases 171-173"
FT misc-structure 186..191
FT /tag- q
FT /note= "theses bases bind to bases 162-167"
FT misc-structure 193..195
FT /tag- r
FT /note= "theses bases bind to bases 158-160"
FT misc-structure 204..208
FT /tag- s
FT /note= "theses bases bind to bases 204-208"
FT WO200202803-A2.
FT 10-JAN-2002.
FT 18-JUN-2001; 2001WC-CH00381.
FT 03-JUL-2000; 2000CH-0001311.
FT (HAFN/) HAFNER W A.
FT (MEN/) MENOUN P.
FT Hafner WA, Menoun P;
FT WPI; 2002-148020/19.
FT
FT Formulating molecular probes, useful for diagnosis and therapy, by
PT analyzing the forms of presentation of targets in organisms
XX
XX Disclosures; Fig 1, 25pp; French.
XX
XX The specification describes a process for formulating and selecting
CC non-contiguous molecular probes which are used for detecting nucleic
CC acids. The process that takes account of the forms of presentation of
CC cellular organisms during the life cycle, or where these forms depend
CC on different chemical, biological and physical states of the organism.
CC The process is a contamination free diagnostic method for direct
CC quantitative and qualitative analysis and gene typing of infectious
CC agents without prior DNA or RNA extraction. The probes enable detection
CC of nucleic acids in sections of living or immobilised, frozen or fixed
CC tissues. Probes of the invention are useful for detecting viral
CC (including oncogenic), bacterial, animal and plant nucleic acid,
CC e.g. in situ hybridization for detecting presence of (pathogenic)
CC microbes and/or determination of genotype, as primers for polymerase
CC chain reaction amplification and therapeutically for blocking
CC replication of pathogens e.g. as ribozymes. Particularly, the probes
CC are used to detect retroviral particles circulating in body fluids or
CC present in cells. The present sequence represents a sequence comprising
CC a consensus endodogmatic sequence/consensus exodogmatic sequence.
XX
XX Sequence 209 BP; 48 A; 59 C; 64 G; 38 U; 0 other;

```

```

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAAGCACCCTATCAGCAGTACACAA 28
DB 78 GCAAGCACCCTATCAGCAGTACACAA 51
RESULT 7
ID AAV70460/C
AC AAV70460;
XX
XX 08-APR-1999 (first entry)
XX
XX Partial sequence of HCV subtype 1b amplicon #86.
XX
XX Nucleic acid detection; nucleic acid characterisation; hybridisation;
XX infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.
XX Hepatitis C virus.
XX
XX WO9850403-A1.
XX
XX 12-NOV-1998.
XX
XX 05-MAY-1998; 98WO-US03194.
XX
XX 03-MAR-1998; 98US-0034205.
XX 05-MAY-1997; 97US-0851588.
XX 19-SEP-1997; 97US-0934097.
XX
XX (THIR-) THIRD WAVE TECHNOLOGIES INC.
XX
XX Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L,
XX Lyamchev VI, Neil BP, Prudent JR;
XX WPI; 1998-610317/51.
XX
XX
XX Detection and characterisation of nucleic acid sequences - by mixing
PT a folded target and one or more probes to form a probe/folded target
PT complex and detecting and characterising the complexes
XX
XX
XX Example 5; Page 172-173; 279pp; English.
XX
XX The invention relates to methods and compositions of detection and
CC characterisation of nucleic acid sequences and sequence changes. One
CC method of detection and characterisation comprises: (a) providing: (1) a
CC folded target having a DNA sequence comprising at least 1 double
CC 1 probe complementary to at least 1 single stranded region; and (ii) at least
CC (b) mixing the target and probes so that the probe hybridises to form a
CC of structure formation in nucleic acid targets; for analysing folded
CC nucleic acids targets; and for analysis of nucleic acid structures. The
CC methods can be used for the detection and characterisation of nucleic
CC acid sequences indicative of an infection, the presence of pathogenic nucleic acid
CC sequences indicative of an infection, the presence of variants or alleles
CC of mammalian genes associated with disease and cancers, and the
CC identification of the source of nucleic acids found in forensic samples,
CC as well as in paternity determinations. The methods allow simultaneous
CC analysis of both strands (e.g. the sense and antisense strands) and are
CC ideal for high-level multiplexing. The products produced are amenable to
CC performed in solution or in the solid phase (e.g. on a solid support).
CC The methods are powerful in that they allow for analysis of longer
CC fragments of nucleic acid than current methodologies. Sequences
CC AAV70453-61 represent partial sequences of different amplicons of
CC hepatitis C virus (HCV) subtypes 1a, 1b, 2c and 3a. These partial
XX sequences are used for identifying the HCV subtypes.
XX
XX Sequence 232 BP; 40 A; 65 C; 76 G; 51 T; 0 other;

```

Query Match 100.0%; Score 28; DB 19; Length 232;
 Best Local Similarity 100.0%; Pred. No. 0.0039;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGCACCCCTATCAGCAGTACCAACA 28
 ||||||||||||||||||||||||||||
 DB 232 GCAGCACCCCTATCAGCAGTACCAACA 205

RESULT 8

ABL46070/c
 ID ABL46070 standard; DNA; 232 BP.

AC ABL46070;

DT 26-APR-2002 (first entry)

DE Hepatitis C virus partial sequence #86 SEQ ID NO:37.

KW Nucleic acid accessible hybridisation site; detection; hybridisation;
 KW Characterisation; identification; nucleic acid structure; diagnosis;

KM PCR primer; probe; ss.

XX Hepatitis C virus.

OS Synthetic.

PN WO200198537-A2.

PD 27-DEC-2001.

PF 15-JUN-2001; 2001WO-US19401.

PR 17-JUN-2000; 2000US-212308P.

PR 15-JUN-2001; 2001US-0212308.

PA (THIR-) THIRD WAVE TECHNOLOGIES INC.

PI Lyemichiev V, Allawi H, Dong F, Neri BP, Vener IT;

DR WPI; 2002-049698/06.

XX Identifying oligonucleotides hybridizing to nucleic acids containing

PT secondary structure, useful in clinical diagnosis, comprises

PT identifying primers that interact with the target to form an extension

PT product under amplification conditions -

PS Example 5; Page 366; 409pp; English.

XX The present invention describes a method for identifying oligonucleotides

CC with desired hybridisation properties to nucleic acid targets containing

CC secondary structure. The method comprises amplifying a target nucleic

CC acid having at least one accessible and one inaccessible site. Primers

CC that form an extension product are identified as the oligonucleotides

CC which can interact with the folded target nucleic acid. Oligonucleotides

CC from the present invention can be used in novel detection methods for

CC clinical diagnostic purposes, including the detection and identification

CC of pathogenic organisms (e.g. HIV). The method allows the ability to

CC rapidly analyse nucleic acid structures. ABL46034 to ABL46367 represent

CC sequences used in the exemplification of the present invention.

XX Sequence 232 BP; 40 A; 65 C; 76 G; 51 T; 0 other;

Query Match 100.0%; Score 28; DB 24; Length 232;

Best Local Similarity 100.0%; Pred. No. 0.0039;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGCACCCCTATCAGCAGTACCAACA 28
 ||||||||||||||||||||||||||||

DB 232 GCAGCACCCCTATCAGCAGTACCAACA 205

RESULT 9

AAV70455/c

ID AAV70455 standard; DNA; 239 BP.

AC AAV70455;

DT 08-APR-1999 (first entry)

DE Partial sequence of HCV subtype 1a amplicon #72.

KW Nucleic acid detection; nucleic acid characterisation; hybridisation;

KW infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.

XX Hepatitis C virus.

OS WO9850403-A1.

PN WO9850403-A1.

PD 12-NOV-1998.

PF 05-MAY-1998; 98WO-US03194.

PR 03-MAR-1998; 98US-0034205.

PR 05-MAY-1997; 97US-0851588.

PR 19-SEP-1997; 97US-0934097.

PA (THIR-) THIRD WAVE TECHNOLOGIES INC.

PI Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L;

PI Lyemichiev VI, Neri BP, Prudent JR;

DR WPI; 1998-610317/51.

PT Detection and characterisation of nucleic acid sequences - by mixing

PT a folded target and one or more probes to form a probe/folded target

PT complex and detecting and characterising the complexes

PS Example 5; Page 171; 279pp; English.

XX The invention relates to methods and compositions of detection and

CC characterisation of nucleic acid sequences and sequence changes. One

CC method of detection and characterisation comprises: (a) providing: (1) a

CC folded target having a DNA sequence comprising at least 1 double

CC stranded region and at least 1 single stranded region; and (11) at least

CC 1 probe complementary to at least a portion of the folded target; and

CC (b) mixing the target and probes so that the probe hybridises to form a

CC probe/folded target complex. Also provided are methods for determination

CC of structure formation in nucleic acid targets; for analysing folded

CC nucleic acids targets; and for analysis of nucleic acid structures. The

CC methods can be used for the detection and characterisation of nucleic

CC acid sequences indicative of an infection, the presence of pathogenic nucleic acid

CC sequences indicative of an infection, the presence of variants or alleles

CC of mammalian genes associated with disease and cancers, and the

CC identification of the source of nucleic acids found in forensic samples,

CC as well as in paternally determinations. The methods allow simultaneous

CC analysis of both strands (e.g. the sense and antisense strands) and are

CC ideal for high-level multiplexing. The products produced are amenable to

CC qualitative, quantitative and positional analysis. The methods may be

CC performed in solution or in the solid phase (e.g. on a solid support).

CC The methods are powerful in that they allow for analysis of longer

CC fragments of nucleic acid than current methodologies. Sequences

CC AAV70453-61 represent partial sequences of different amplicons of

CC hepatitis C virus (HCV) subtypes 1a, 1b, 2c and 3a. These partial

CC sequences are used for identifying the HCV subtypes.

XX Sequence 239 BP; 45 A; 63 C; 78 G; 53 T; 0 other;

Query Match 100.0%; Score 28; DB 19; Length 239;

Best Local Similarity 100.0%; Pred. No. 0.0039;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGCACCCCTATCAGCAGTACCAACA 28
 ||||||||||||||||||||||||||||

DB 239 GCAGCACCCCTATCAGCAGTACCAACA 212

RESULT 10
AAV70459/c
ID AAV70459 standard; DNA: 239 BP.
XX
XX AAV70459;
XX
XX 08-APR-1999 (first entry)
XX
XX Partial sequence of HCV subtype 1a amplicon #85.
XX
XX Nucleic acid detection; nucleic acid characterisation; hybridisation;
XX infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.
XX
XX Hepatitis C virus.
XX
XX WO9850403-A1.
XX
XX 12-NOV-1998.
XX
XX 05-MAY-1998; 98WO-US03194.
XX
XX 03-MAR-1998; 98US-0034205.
XX 05-MAY-1997; 97US-0851588.
XX 19-SEP-1997; 97US-0934097.
XX
XX (THIR-) THIRD WAVE TECHNOLOGIES INC.
XX
XX Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L;
XX Lyamichev VI, Neri BP, Prudent JR;
XX WPI; 1998-610317/51.
XX
XX Detection and characterisation of nucleic acid sequences - by mixing
XX a folded target and one or more probes to form a probe/folded target
XX complex and detecting and characterising the complexes
XX
XX Example 5; Page 172; 279pp; English.
XX
XX The invention relates to methods and compositions of detection and
XX characterisation of nucleic acid sequences and sequence changes. One
XX method of detection and characterisation comprises: (a) providing: (1) a
XX folded target having a DNA sequence comprising at least 1 double
XX stranded region and at least 1 single stranded region; and (11) at least
XX 1 probe complementary to at least a portion of the folded target; and
XX (b) mixing the target and probes so that the probe hybridises to form a
XX probe/folded target complex. Also provided are methods for determination
XX of structure formation in nucleic acid targets; for analysing folded
XX nucleic acids targets; and for analysis of nucleic acid structures. The
XX methods can be used for the detection and characterisation of nucleic
XX acid sequences to detect the presence of pathogenic nucleic acid
XX sequences indicative of an infection, the presence of variants or alleles
XX of mammalian genes associated with disease and cancers, and the
XX identification of the source of nucleic acids found in forensic samples,
XX as well as in paternity determinations. The methods allow simultaneous
XX analysis of both strands (e.g. the sense and antisense strands) and are
XX ideal for high-level multiplexing. The products produced are amenable to
XX qualitative, quantitative and positional analysis. The methods may be
XX performed in solution or in the solid phase (e.g. on a solid support).
XX The methods are powerful in that they allow for analysis of longer
XX fragments of nucleic acid than current methodologies. Sequences
XX AAV70453-61 represent partial sequences of different amplicons of
XX hepatitis C virus (HCV) subtypes 1a, 1b, 2c and 3a. These partial
XX sequences are used for identifying the HCV subtypes.
XX
XX Sequence 239 BP; 45 A; 64 C; 77 G; 53 T; 0 other;
SO
Query Match 100.0%; Score 28; DB 19; Length 239;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Caps 0;

RESULT 11
ABL46065/c
ID ABL46065 standard; DNA: 239 BP.
XX
XX ABL46065;
XX
XX 26-APR-2002 (first entry)
XX
XX Hepatitis C virus partial sequence #72 SEQ ID NO:32.
XX
XX Nucleic acid accessible hybridisation site; detection; hybridisation;
XX characterisation; identification; nucleic acid structure; diagnosis;
XX PCR primer; probe; ss.
XX
XX Hepatitis C virus.
XX
XX Synthetic.
XX
XX WO200198537-A2.
XX
XX 27-DEC-2001.
XX
XX 15-JUN-2001; 2001WO-US19401.
XX
XX 17-JUN-2000; 2000US-212308P.
XX 15-JUN-2001; 2001US-0212308.
XX
XX (THIR-) THIRD WAVE TECHNOLOGIES INC.
XX
XX Lyamichev V, Allawi H, Dong F, Neri BP, Vener IT;
XX WPI; 2002-049698/06.
XX
XX Identifying oligonucleotides hybridizing to nucleic acids containing
XX secondary structure, useful in clinical diagnosis, comprises
XX identifying primers that interact with the target to form an extension
XX product under amplification conditions -
XX
XX Example 5; Page 365; 409pp; English.
XX
XX The present invention describes a method for identifying oligonucleotides
XX with desired hybridisation properties to nucleic acid targets containing
XX secondary structure. The method comprises amplifying a target nucleic
XX acid having at least one accessible and one inaccessible site. Primers
XX that form an extension product are identified as the oligonucleotides
XX which can interact with the folded target nucleic acid. Oligonucleotides
XX from the present invention can be used in novel detection methods for
XX clinical diagnostic purposes, including the detection and identification
XX of pathogenic organisms (e.g. HIV). The method allows the ability to
XX rapidly analyse nucleic acid structures. ABL46034 to ABL46367 represent
XX sequences used in the exemplification of the present invention.
XX
XX Sequence 239 BP; 45 A; 63 C; 78 G; 53 T; 0 other;
SO
Query Match 100.0%; Score 28; DB 24; Length 239;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Caps 0;

RESULT 12
ABL46069/c
ID ABL46069 standard; DNA: 239 BP.
XX
XX ABL46069;
XX
XX 26-APR-2002 (first entry)
XX
XX Hepatitis C virus partial sequence #85 SEQ ID NO:36.
XX

XX Nucleic acid accessible hybridisation site; detection; hybridisation;
 KW characterisation; identification; nucleic acid structure; diagnosis;
 KM PCR primer; probe; ss.
 XX
 XX Hepatitis C virus.
 OS Synthetic.
 XX
 XX WO200198537-A2.
 PN
 XX
 PD 27-DEC-2001.
 XX
 XX 15-JUN-2001; 2001WO-US19401.
 PE
 XX 17-JUN-2000; 2000US-212308P.
 PR
 XX 15-JUN-2001; 2001US-0212308.
 PR
 XX
 PA (THIR-) THIRD WAVE TECHNOLOGIES INC.
 XX
 XX Lyamchev V, Allawi H, Dong F, Neri BP, Vener IT;
 PI
 XX WPI; 2002-049698/06.
 DR
 XX
 XX Identifying oligonucleotides hybridizing to nucleic acids containing
 PT secondary structure, useful in clinical diagnosis, comprises
 PT identifying primers that interact with the target to form an extension
 PT product under amplification conditions -
 XX
 XX Example 5; Page 366; 409pp; English.
 PS
 XX The present invention describes a method for identifying oligonucleotides
 CC with desired hybridisation properties to nucleic acid targets containing
 CC secondary structure. The method comprises amplifying a target nucleic
 CC acid having at least one accessible and one inaccessible site. Primers
 CC that form an extension product are identified as the oligonucleotides
 CC which can interact with the folded target nucleic acid. Oligonucleotides
 CC from the present invention can be used in novel detection methods for
 CC clinical diagnostic purposes, including the detection and identification
 CC of pathogenic organisms (e.g. HIV). The method allows the ability to
 CC rapidly analyse nucleic acid structures. ABL46034 to ABL46367 represent
 CC sequences used in the exemplification of the present invention.
 CC
 SQ Sequence 239 BP; 45 A; 64 C; 77 G; 53 T; 0 other;
 Query Match 100.0%; Score 28; DB 24; Length 239;
 Best Local Similarity 100.0%; Pred. No. 0.0039;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GCAGCACCCTATCAGGAGTACACAA 28
 DB 239 GCAGCACCCTATCAGGAGTACACAA 212
 RESULT 13
 AAV70457/C
 ID AAV70457 standard; DNA; 240 BP.
 XX
 AC AAV70457;
 XX
 DT 08-APR-1999 (first entry)
 XX
 DE Partial sequence of HCV subtype 2b amplicon #74.
 XX
 KM Nucleic acid detection; nucleic acid characterisation; hybridisation;
 KM infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.
 XX
 OS Hepatitis C virus.
 XX
 PN WO9850403-A1.
 XX
 PD 12-NOV-1998.
 XX
 PF 05-MAY-1998; 98WO-US03194.

XX
 PR 03-MAR-1998; 98US-0034205.
 PR 05-MAY-1997; 97US-0851588.
 PR 19-SEP-1997; 97US-0934097.
 XX
 XX (THIR-) THIRD WAVE TECHNOLOGIES INC.
 PA
 PI Anderson TR, Brow MAD, Dahlberg JE, Dong F, Fors L;
 PI Lyamchev VI, Neri BP, Prudent JR;
 XX
 XX WPI; 1998-610317/51.
 DR
 XX
 XX Detection and characterisation of nucleic acid sequences - by mixing
 PT a folded target and one or more probes to form a probe/folded target
 PT complex and detecting and characterising the complexes
 PT
 XX
 XX Example 5; Page 171-172; 279pp; English.
 PS
 XX The invention relates to methods and compositions of detection and
 CC characterisation of nucleic acid sequences and sequence changes. One
 CC method of detection and characterisation comprises: (a) providing: (1) a
 CC folded target having a DNA sequence comprising at least 1 double
 CC stranded region and at least 1 single stranded region; and (1i) at least
 CC 1 probe complementary to at least a portion of the folded target; and
 CC (b) mixing the target and probes so that the probe hybridises to form a
 CC probe/folded target complex. Also provided are methods for determination
 CC of structure formation in nucleic acid targets; for analysing folded
 CC nucleic acids targets; and for analysis of nucleic acid structures. The
 CC methods can be used for the detection and characterisation of nucleic
 CC acid sequences to detect the presence of pathogenic nucleic acid
 CC sequences indicative of an infection, the presence of variants or alleles
 CC of mammalian genes associated with disease and cancers, and the
 CC identification of the source of nucleic acids found in forensic samples,
 CC as well as in paternity determinations. The methods allow simultaneous
 CC analysis of both strands (e.g. the sense and antisense strands) and are
 CC ideal for high-level multiplexing. The products produced are amenable to
 CC qualitative, quantitative and positional analysis. The methods may be
 CC performed in solution or in the solid phase (e.g. on a solid support).
 CC The methods are powerful in that they allow for analysis of longer
 CC fragments of nucleic acid than current methodologies. Sequences
 CC AAV70453-61 represent partial sequences of different amplicons of
 CC hepatitis C virus (HCV) subtypes 1a, 1b, 2c and 3a. These partial
 CC sequences are used for identifying the HCV subtypes.
 CC
 SQ Sequence 240 BP; 47 A; 64 C; 74 G; 55 T; 0 other;
 Query Match 100.0%; Score 28; DB 19; Length 240;
 Best Local Similarity 100.0%; Pred. No. 0.0039;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GCAGCACCCTATCAGGAGTACACAA 28
 DB 240 GCAGCACCCTATCAGGAGTACACAA 213
 RESULT 14
 AAV70456/C
 ID AAV70456 standard; DNA; 240 BP.
 XX
 AC AAV70456;
 XX
 DT 08-APR-1999 (first entry)
 XX
 DE Partial sequence of HCV subtype 1a amplicon #73.
 XX
 KM Nucleic acid detection; nucleic acid characterisation; hybridisation;
 KM infection; disease; cancer; forensic; paternity; multiplexing; HCV; ds.
 XX
 OS Hepatitis C virus.
 XX
 PN WO9850403-A1.
 XX
 PD 12-NOV-1998.

```

XX PF 05-MAY-1998; 98WO-US03194.
XX PR 03-MAR-1998; 98US-0034205.
XX PR 05-MAY-1997; 97US-0851588.
XX PR 19-SEP-1997; 97US-0934097.
XX PA (THIR-) THIRD WAVE TECHNOLOGIES INC.
XX PI Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L;
XX PI Lyamchev VI, Neri BP, Prudent JR;
XX DR WPI; 1998-610317/51.
XX PT Detection and characterisation of nucleic acid sequences - by mixing
XX PT a folded target and one or more probes to form a probe/folded target
XX PT complex and detecting and characterising the complexes
XX PS Example 5; Page 171; 279pp; English.
XX CC The invention relates to methods and compositions of detection and
XX CC characterisation of nucleic acid sequences and sequence changes. One
XX CC method of detection and characterisation comprises: (a) providing: (1) a
XX CC folded target having a DNA sequence comprising at least 1 double
XX CC stranded region and at least 1 single stranded region; and (11) at least
XX CC 1 probe complementary to at least a portion of the folded target; and
XX CC (b) mixing the target and probes so that the probe hybridises to form a
XX CC probe /folded target complex. Also provided are methods for determination
XX CC of structure formation in nucleic acid targets; for analysing folded
XX CC nucleic acids targets; and for analysis of nucleic acid structures. The
XX CC methods can be used for the detection and characterisation of nucleic
XX CC acid sequences to detect the presence of pathogenic nucleic acid
XX CC sequences indicative of an infection, the presence of variants or alleles
XX CC of mammalian genes associated with disease and cancers, and the
XX CC identification of the source of nucleic acids found in forensic samples,
XX CC as well as in paternity determinations. The methods allow simultaneous
XX CC analysis of both strands (e.g. the sense and antisense strands) and are
XX CC ideal for high-level multiplexing. The products produced are amenable to
XX CC qualitative, quantitative and positional analysis. The methods may be
XX CC performed in solution or in the solid phase (e.g. on a solid support).
XX CC The methods are powerful in that they allow for analysis of longer
XX CC fragments of nucleic acid than current methodologies. Sequences
XX CC AAV70453-61 represent partial sequences of different amplicons of
XX CC hepatitis C virus (HCV) subtypes 1a, 1b, 2c and 3a. These partial
XX CC sequences are used for identifying the HCV subtypes.
XX SQ Sequence 240 BP; 47 A; 63 C; 78 G; 52 T; 0 other;
XX
Query Match 100.0%; Score 28; DB 19; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GCAGACCCCTATCAGGACGATACCAAA 28
DB 240 GCAGACCCCTATCAGGACGATACCAAA 213

```

```

XX PD 12-NOV-1998.
XX PF 05-MAY-1998; 98WO-US03194.
XX PR 03-MAR-1998; 98US-0034205.
XX PR 05-MAY-1997; 97US-0851588.
XX PR 19-SEP-1997; 97US-0934097.
XX PA (THIR-) THIRD WAVE TECHNOLOGIES INC.
XX PI Anderson TA, Brow MAD, Dahlberg JE, Dong F, Fors L;
XX PI Lyamchev VI, Neri BP, Prudent JR;
XX DR WPI; 1998-610317/51.
XX PT Detection and characterisation of nucleic acid sequences - by mixing
XX PT a folded target and one or more probes to form a probe/folded target
XX PT complex and detecting and characterising the complexes
XX PS Example 5; Page 172; 279pp; English.
XX CC The invention relates to methods and compositions of detection and
XX CC characterisation of nucleic acid sequences and sequence changes. One
XX CC method of detection and characterisation comprises: (a) providing: (1) a
XX CC folded target having a DNA sequence comprising at least 1 double
XX CC stranded region and at least 1 single stranded region; and (11) at least
XX CC 1 probe complementary to at least a portion of the folded target; and
XX CC (b) mixing the target and probes so that the probe hybridises to form a
XX CC probe /folded target complex. Also provided are methods for determination
XX CC of structure formation in nucleic acid targets; for analysing folded
XX CC nucleic acids targets; and for analysis of nucleic acid structures. The
XX CC methods can be used for the detection and characterisation of nucleic
XX CC acid sequences indicative of an infection, the presence of variants or alleles
XX CC of mammalian genes associated with disease and cancers, and the
XX CC identification of the source of nucleic acids found in forensic samples,
XX CC as well as in paternity determinations. The methods allow simultaneous
XX CC analysis of both strands (e.g. the sense and antisense strands) and are
XX CC ideal for high-level multiplexing. The products produced are amenable to
XX CC qualitative, quantitative and positional analysis. The methods may be
XX CC performed in solution or in the solid phase (e.g. on a solid support).
XX CC The methods are powerful in that they allow for analysis of longer
XX CC fragments of nucleic acid than current methodologies. Sequences
XX CC AAV70453-61 represent partial sequences of different amplicons of
XX CC hepatitis C virus (HCV) subtypes 1a, 1b, 2c and 3a. These partial
XX CC sequences are used for identifying the HCV subtypes.
XX SQ Sequence 240 BP; 45 A; 66 C; 79 G; 50 T; 0 other;
XX
Query Match 100.0%; Score 28; DB 19; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GCAGACCCCTATCAGGACGATACCAAA 28
DB 240 GCAGACCCCTATCAGGACGATACCAAA 213

```

Search completed: July 10, 2003, 19:52:42
 Job time : 60.0831 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 ; Search time 12.5213 Seconds
(without alignments)
685.785 Million cell updates/sec

Title: US-10-087-631B-7

Perfect score: 28

Sequence: 1 gcaagcaccctatcagcagcagcacaac 28

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	100.0	28	2	US-08-738-928-2
2	28	100.0	28	3	US-09-039-866-4
3	28	100.0	47	4	US-09-547-757-1
4	28	100.0	57	1	US-08-356-287-36
5	28	100.0	57	5	PCR-US93-04863-36
6	28	100.0	232	4	US-09-034-205-37
7	28	100.0	232	4	US-08-851-588-37
8	28	100.0	232	4	US-08-934-097A-37
9	28	100.0	232	4	US-09-677-218B-37
10	28	100.0	232	4	US-09-677-218B-37
11	28	100.0	233	4	US-09-034-205-32
12	28	100.0	233	4	US-09-034-205-36
13	28	100.0	233	4	US-08-934-097A-32
14	28	100.0	233	4	US-08-934-097A-36
15	28	100.0	233	4	US-08-851-588-32
16	28	100.0	233	4	US-08-851-588-36
17	28	100.0	233	4	US-09-677-218B-32
18	28	100.0	233	4	US-09-677-218B-36
19	28	100.0	233	4	US-09-677-192-32
20	28	100.0	239	4	US-09-677-192-36
21	28	100.0	240	4	US-09-034-205-33
22	28	100.0	240	4	US-09-034-205-34
23	28	100.0	240	4	US-09-034-205-35
24	28	100.0	240	4	US-08-934-097A-33
25	28	100.0	240	4	US-08-934-097A-34
26	28	100.0	240	4	US-08-934-097A-35
27	28	100.0	240	4	US-08-851-588-33

28	100.0	240	4	US-08-851-588-34	Sequence 34, Appl
28	100.0	240	4	US-08-851-588-35	Sequence 35, Appl
28	100.0	240	4	US-09-677-218B-33	Sequence 33, Appl
28	100.0	240	4	US-09-677-218B-34	Sequence 34, Appl
28	100.0	240	4	US-09-677-192-33	Sequence 35, Appl
28	100.0	240	4	US-09-677-192-34	Sequence 36, Appl
28	100.0	240	4	US-09-677-192-35	Sequence 37, Appl
28	100.0	244	4	US-09-034-205-26	Sequence 26, Appl
28	100.0	244	4	US-09-034-205-27	Sequence 27, Appl
28	100.0	244	4	US-09-034-205-28	Sequence 28, Appl
28	100.0	244	4	US-09-034-205-29	Sequence 29, Appl
28	100.0	244	4	US-09-034-205-31	Sequence 31, Appl
28	100.0	244	4	US-08-934-097A-26	Sequence 26, Appl
28	100.0	244	4	US-08-934-097A-27	Sequence 27, Appl
28	100.0	244	4	US-08-934-097A-28	Sequence 28, Appl
28	100.0	244	4	US-08-934-097A-29	Sequence 29, Appl
28	100.0	244	4	US-08-934-097A-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-08-738-928-2
Sequence 2, Application US/08738928
Patent No. 5837442

GENERAL INFORMATION:

APPLICANT: Tsang, Sue Y.

TITLE OF INVENTION: Oligonucleotide Primers for Amplifying

TITLE OF INVENTION: HCV Nucleic Acid

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: U.S.A.

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/738,928

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Petry, Douglas A.

REGISTRATION NUMBER: 35,321

REFERENCE/DOCKET NUMBER: 9263

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 814-2974

TELEFAX: (510) 814-2977

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 28 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-738-928-2

Query Match 100.0%, Score 28; DB 2; Length 28;

Best Local Similarity 100.0%, Pred. No. 0.00011;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGCAGCAGCACA 28
DB 1 GCAAGCACCCTATCAGCAGCAGCACA 28

RESULT 2

US-09-039-866-4
Sequence 4, Application US/09039866
Patent No. 6001611

GENERAL INFORMATION:

APPLICANT: Will, Stephen G.
TITLE OF INVENTION: MODIFIED NUCLEIC ACID AMPLIFICATION
TITLE OF INVENTION: PRIMERS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Roche Molecular Systems
STREET: 1080 U.S. Highway 202
CITY: Branchburg
STATE: New Jersey
COUNTRY: United States
ZIP: 08876

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/039,866
FILING DATE:
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Pety, Douglas A.
REGISTRATION NUMBER: 35,321
REFERENCE/DOCKET NUMBER: 1023P
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 28 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-09-039-866-4

Query Match
Best Local Similarity 100.0%; Score 28; DB 3; Length 28;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAAGCACCCTATCAGGAGTACCA 28
DB 1 GCAAGCACCCTATCAGGAGTACCA 28

RESULT 3

US-09-547-757-1/c
Sequence 1, Application US/09547757
Patent No. 6368801

GENERAL INFORMATION:

APPLICANT: Faruqi, A. Fawad
TITLE OF INVENTION: Detection and Amplification of RNA using
FILE REFERENCE: MSI 102
CURRENT APPLICATION NUMBER: US/09/547,757
CURRENT FILING DATE: 2000-06-21
NUMBER OF SEQ ID NOS: 1
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1
LENGTH: 47
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic

US-09-547-757-1

Query Match

Best Local Similarity 100.0%; Score 28; DB 4; Length 47;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY

DB 1 GCAAGCACCCTATCAGGAGTACCA 28
36 GCAAGCACCCTATCAGGAGTACCA 9

RESULT 4

US-08-356-287-36/c
Sequence 36, Application US/08356287
Patent No. 5686272

GENERAL INFORMATION:

APPLICANT: Ronald L. Marshall
APPLICANT: John J. Carrino
TITLE OF INVENTION: AMPLIFICATION OF RNA SEQUENCES USING
TITLE OF INVENTION: THE LIGASE CHAIN REACTION
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy diskette
COMPUTER: Macintosh
OPERATING SYSTEM: System 7.0.1
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356,287
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/891,543
FILING DATE: 29 MAY 1992
ATTORNEY/AGENT INFORMATION:

NAME: Paul D. Yassier
REGISTRATION NUMBER: 37,477
REFERENCE/DOCKET NUMBER: 5172, US, P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-2341
TELEFAX: 708-938-2623

INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 57
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA

US-08-356-287-36

Query Match
Best Local Similarity 100.0%; Score 28; DB 1; Length 57;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAAGCACCCTATCAGGAGTACCA 28
DB 54 GCAAGCACCCTATCAGGAGTACCA 27

RESULT 5

PCT-US93-04863-36/c

Sequence 36, Application PC/TUS9304863

GENERAL INFORMATION:

APPLICANT: Ronald L. Marshall
APPLICANT: John J. Carrino
APPLICANT: Joann C. Sustachek
TITLE OF INVENTION: AMPLIFICATION OF RNA SEQUENCES
TITLE OF INVENTION: USING THE LIGASE CHAIN REACTION
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories

STREET: One Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy diskette
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04863
FILING DATE: 19930524
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/891,543
FILING DATE: 29 MAY 1992
ATTORNEY/AGENT INFORMATION:
NAME: Thomas D. Bralhard
REGISTRATION NUMBER: 32,459
REFERENCE/DOCKET NUMBER: 5172.PC.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-4884
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 57
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA
PCT-US93-04863-36
Query Match 100.0%; Score 28; DB 5; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCACGACCCCTATCAGCGATACACAA 28
DB 54 GCACGACCCCTATCAGCGATACACAA 27
RESULT 6
US-09-034-205-37/C
Sequence 37, Application US/09034205
Patent No. 6194149
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Nerli, Bruce P.
TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING
STRUCTURE-BRIDGING OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,205
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: 5172.PC.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-4884
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 57
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA
PCT-US93-04863-36

REFERENCE/DOCKET NUMBER: FORS-03268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-034-205-37
Query Match 100.0%; Score 28; DB 4; Length 232;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCACGACCCCTATCAGCGATACACAA 28
DB 232 GCACGACCCCTATCAGCGATACACAA 205
RESULT 7
US-08-934-097A-37/C
Sequence 37, Application US/08934097A
Patent No. 6210880
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Nerli, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
STRUCTURE PROBING WITH STRUCTURE-BRIDGING
OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/934,097A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-08-934-097A-37
Query Match 100.0%; Score 28; DB 4; Length 232;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCTATCAGCAGTACCACAA 28
|||||
Db 232 GCAAGCACCTATCAGCAGTACCACAA 205

RESULT 8
US-08-851-588-37/C
Sequence 37, Application US/08851588
Patent No. 6214545

GENERAL INFORMATION:

APPLICANT: Doog, Fang
APPLICANT: Lyamichev, Victor I.
APPLICANT: Prudent, James R.
APPLICANT: Dahlberg, James E.
APPLICANT: Fors, Lance
TITLE OF INVENTION: Polymorphism Analysis by Nucleic Acid
TITLE OF INVENTION: Structure Probing
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,588
FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02777
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:
LENGTH: 232 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-08-851-588-37

Query Match 100.0%; Score 28; DB 4; Length 232;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCTATCAGCAGTACCACAA 28
|||||
Db 232 GCAAGCACCTATCAGCAGTACCACAA 205

RESULT 9
US-09-677-218B-37/C

Sequence 37, Application US/09677218B
Patent No. 6355437

GENERAL INFORMATION:

APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Neil, Bruce P.

TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING

STRUCTURE-BRIDGING OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/677,218B
FILING DATE: 02-Oct-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/034,205

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-03268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:

LENGTH: 232 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

US-09-677-218B-37
SEQUENCE DESCRIPTION: SEQ ID NO: 37:

Query Match 100.0%; Score 28; DB 4; Length 232;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCTATCAGCAGTACCACAA 28
|||||
Db 232 GCAAGCACCTATCAGCAGTACCACAA 205

RESULT 10

US-09-677-192-37/C

Sequence 37, Application US/09677192
Patent No. 6358691

GENERAL INFORMATION:

APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Neil, Bruce P.

TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING STRUCTURE-BRIDGING

FILE REFERENCE: FORS-04708
CURRENT APPLICATION NUMBER: US/09/677,192
CURRENT FILING DATE: 2000-10-02

PRIOR APPLICATION NUMBER: 09/034,205
PRIOR FILING DATE: 1998-03-03

NUMBER OF SEQ ID NOS: 68
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 37

LENGTH: 232

TYPE: DNA

ORGANISM: Hepatitis C virus
US-09-677-192-37

Query Match 100.0%; Score 28; DB 4; Length 232;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Fri Jul 11 15:29:37 2003

QY 1 GCAAGCACCCCTATCAGGAGTACCA 28
DB 232 GCAAGCACCCCTATCAGGAGTACCA 205

RESULT 11
US-09-034-205-32/c
Sequence 32, Application US/09034205
Patent No. 6194149
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Neel, Bruce P.
TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING
STRUCTURE-BRIDGING OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,205
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamlin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-03268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 239 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-034-205-32

Query Match 100.0%; Score 28; DB 4; Length 239;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCCTATCAGGAGTACCA 28
DB 239 GCAAGCACCCCTATCAGGAGTACCA 212

RESULT 12
US-09-034-205-36/c
Sequence 36, Application US/09034205
Patent No. 6194149
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Neel, Bruce P.
TITLE OF INVENTION: TARGET-DEPENDENT REACTIONS USING
STRUCTURE-BRIDGING OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,205
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamlin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-03268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 239 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-034-205-36

Query Match 100.0%; Score 28; DB 4; Length 239;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCCTATCAGGAGTACCA 28
DB 239 GCAAGCACCCCTATCAGGAGTACCA 212

RESULT 13
US-08-934-097A-32/c
Sequence 32, Application US/08934097A
Patent No. 6210880
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Fors, Lance
APPLICANT: Neel, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
STRUCTURE PROBING WITH STRUCTURE-BRIDGING
OLIGONUCLEOTIDES
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/934,097A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamlin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 239 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

US-08-934-097A-32

OY 1 GCAGACCCCTATCAGCAGTACCACAA 28
DB 239 GCAGACCCCTATCAGCAGTACCACAA 212

RESULT 14

US-08-934-097A-36/C

Sequence 36, Application US/08934097A

Patent No. 6210880

GENERAL INFORMATION:

APPLICANT: Lyamichev, Victor I.

APPLICANT: Brow, Mary Ann D.

APPLICANT: Fors, Lance

APPLICANT: Neri, Bruce P.

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid

TITLE OF INVENTION: Structure Probing With Structure-Bridging

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESS: MEDLEN & CARROLL, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/934,097A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Macknight, Kamelin T.

REGISTRATION NUMBER: 38,230

REFERENCE/DOCKET NUMBER: FORS-02980

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 36:

SEQUENCE CHARACTERISTICS:

LENGTH: 239 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

US-08-934-097A-36

OY 1 GCAGACCCCTATCAGCAGTACCACAA 28

DB 239 GCAGACCCCTATCAGCAGTACCACAA 212

Query Match 100.0%; Score 28; DB 4; Length 239;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB

239 GCAGACCCCTATCAGCAGTACCACAA 212

RESULT 15

US-08-851-588-32/C

Sequence 32, Application US/08851588

Patent No. 6214545

GENERAL INFORMATION:

APPLICANT: Dong, Fang

APPLICANT: Lyamichev, Victor I.

APPLICANT: Prudent, James R.

APPLICANT: Dahlberg, James E.

APPLICANT: Fors, Lance

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid

TITLE OF INVENTION: Structure Probing

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESS: MEDLEN & CARROLL, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/851,588

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: FORS-02777

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 239 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

US-08-851-588-32

OY 1 GCAGACCCCTATCAGCAGTACCACAA 28
DB 239 GCAGACCCCTATCAGCAGTACCACAA 212

Search completed: July 10, 2003, 20:27:39
Job time: 12.5213 secs

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:53:16 ; Search time 76.6382 Seconds
(without alignments)
576.891 Million cell updates/sec

Title: US-10-087-631B-7

Sequence: 1 gcaagcaccctatcagcgagtlaccacaa 28

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: : Published_Applications_NA:*

```

1: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCONB.seq.*
5: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCONB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US07_PUBCONB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCONB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCONB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCONB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCONB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	28	100.0	28	9	US-10-087-631B-7	Sequence 7, Appl1
2	28	100.0	60	9	US-09-870-939-1	Sequence 1, Appl1
3	28	100.0	232	10	US-09-825-574-37	Sequence 37, App
4	28	100.0	239	10	US-09-825-574-32	Sequence 32, App
5	28	100.0	239	10	US-09-825-574-36	Sequence 36, App
6	28	100.0	240	10	US-09-825-574-33	Sequence 34, App
7	28	100.0	240	10	US-09-825-574-34	Sequence 33, App
8	28	100.0	240	10	US-09-825-574-35	Sequence 35, App
9	28	100.0	241	9	US-10-087-631B-10	Sequence 10, Appl1
10	28	100.0	241	9	US-10-087-631B-13	Sequence 13, Appl1
11	28	100.0	242	9	US-10-087-631B-15	Sequence 15, Appl1
12	28	100.0	244	10	US-09-825-574-26	Sequence 26, App
13	28	100.0	244	10	US-09-825-574-27	Sequence 27, App
14	28	100.0	244	10	US-09-825-574-28	Sequence 28, App
15	28	100.0	244	10	US-09-825-574-29	Sequence 29, App
16	28	100.0	244	10	US-09-825-574-31	Sequence 31, App
17	28	100.0	281	9	US-09-940-925A-123	Sequence 123, App
18	28	100.0	281	9	US-09-940-925A-125	Sequence 125, App
19	28	100.0	281	9	US-09-940-925A-126	Sequence 126, App

20	28	100.0	281	9	US-09-940-9255A-128	Sequence 128, App
21	28	100.0	281	9	US-09-940-9255A-129	Sequence 129, App
22	28	100.0	281	9	US-09-940-9255A-131	Sequence 131, App
23	28	100.0	281	9	US-09-940-9255A-132	Sequence 132, App
24	28	100.0	281	9	US-09-941-1933A-125	Sequence 125, App
25	28	100.0	281	9	US-09-941-1933A-126	Sequence 126, App
26	28	100.0	281	9	US-09-941-1933A-128	Sequence 128, App
27	28	100.0	281	9	US-09-941-1933A-129	Sequence 129, App
28	28	100.0	281	9	US-09-941-1933A-131	Sequence 131, App
29	28	100.0	281	9	US-09-941-1933A-132	Sequence 132, App
30	28	100.0	281	9	US-09-941-1933A-134	Sequence 134, App
31	28	100.0	282	9	US-09-940-9255A-124	Sequence 124, App
32	28	100.0	282	9	US-09-940-9255A-130	Sequence 130, App
33	28	100.0	282	9	US-09-941-1933A-124	Sequence 124, App
34	28	100.0	282	9	US-09-941-1933A-130	Sequence 130, App
35	28	100.0	286	10	US-09-825-574-21	Sequence 21, App
36	28	100.0	289	10	US-09-825-574-22	Sequence 22, App
37	28	100.0	289	10	US-09-825-574-22	Sequence 22, App
38	28	100.0	289	10	US-09-825-574-23	Sequence 23, App
39	28	100.0	298	10	US-09-345-761-6	Sequence 6, App
40	28	100.0	315	10	US-09-345-761-6	Sequence 6, App
41	28	100.0	341	9	US-09-814-357-3	Sequence 3, App
42	28	100.0	341	9	US-10-259-275-35	Sequence 35, App
43	28	100.0	341	10	US-09-814-295-14	Sequence 44, App
44	28	100.0	347	9	US-10-137-295-1	Sequence 1, App
45	28	100.0	366	9	US-09-992-160-48	Sequence 48, App

ALIGNMENTS

Sequence 48, Appl

RESULT 1
US-10-087-631B-7

; Sequence /, Application US/10087631E
; Publication No. US20030054372A1
GENERAL INFORMATION

GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE

FILE REFERENCE: 1803-335-999

CURRENT APPLICATION NUMBER: US/10/0
CURRENT FILING DATE: 2002-03-01

```

; NUMBER OF SEQ ID NOS:.17
; SOFTWARE: PatentIn version 3.1

```

```

; SEQ ID NO 7
; LENGTH: 28

```

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

```

OTHER INFORMATION:	Description of
OTHER INFORMATION:	sequence

US-10-087-631B-7

Query Match	100.0%;	Score
Best Local Similarity	100.0%;	Pre

Matches	28;	Conservative	0;	M
---------	-----	--------------	----	---

1 GCAAGCACCCTATCAGGCAGTACC
1 |||||

DD 1 GCATAGCACCCTATTCAGGCAATACC

RESULT 2
US-09-870-939-1/c

; Sequence 1, Application US/09870939
; Publication No. US20020192650A1

```

; GENERAL INFORMATION:
; APPLICANT: AMORESE, DOUGLAS A.

```

APPLICANT: SHANNON, KAREN W.
APPLICANT: COLLINS, PATRICK J.

```

; APPLICANT: WOLBER, PAUL K.
; TITLE OF INVENTION: COMPOSITE ARRAY
; REFERENCE: 10010201-1

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Fri Jul 11 15:29:38 2003

us-10-087-631b-7.rnpb

CURRENT APPLICATION NUMBER: US/09/870,939
NUMBER OF FILING DATE: 2001-10-12
SOFTWARE: SEQ ID NOS: 1
SEQ ID NO: 1
LENGTH: 60
TYPE: DNA
ORGANISM: Hepatitis C virus
US-09-870-939-1

Query Match

Best Local

Matches 28; Similarity 100.0%; Score 28; DB 9; Length 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 GCAGCACCCCTATCAGCAGTACACCA 28
57 GCAGCACCCCTATCAGCAGTACACCA 28
Indels 0; Gaps 0;

RESULT 3
US-09-825-574-37/c
Sequence 37, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamichiev, Victor I.
Brow, Mary Ann D.
Neil, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
CORRESPONDENCE ADDRESS: 51
ADDRESS: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER: IBM PC compatible
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: Patentin Release #1.0, Version #1.30
FILING DATE: 03-Apr-2001
CLASSIFICATION: US/09/825,574
PRIOR APPLICATION DATA: <Unknown>
APPLICATION NUMBER: 08/934,097
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION/DOCKET NUMBER: 38,230
TELECOMMUNICATION INFORMATION: FONS-02980
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
SEQUENCE FOR SEQ ID NO: 37:
LENGTH: 232
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: other nucleic acid
TOPOLOGY: linear
SEQUENCE DESCRIPTION: /desc = "DNA"

US-09-825-574-37/c
Sequence 37, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamichiev, Victor I.
Brow, Mary Ann D.
Neil, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
CORRESPONDENCE ADDRESS: 51
ADDRESS: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER: IBM PC compatible
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: Patentin Release #1.0, Version #1.30
FILING DATE: 03-Apr-2001
CLASSIFICATION: US/09/825,574
PRIOR APPLICATION DATA: <Unknown>
APPLICATION NUMBER: 08/934,097
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION/DOCKET NUMBER: 38,230
TELECOMMUNICATION INFORMATION: FONS-02980
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
SEQUENCE FOR SEQ ID NO: 37:
LENGTH: 232
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: other nucleic acid
TOPOLOGY: linear
SEQUENCE DESCRIPTION: /desc = "DNA"

US-09-825-574-37/c
Sequence 37, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamichiev, Victor I.
Brow, Mary Ann D.
Neil, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
CORRESPONDENCE ADDRESS: 51
ADDRESS: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER: IBM PC compatible
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: Patentin Release #1.0, Version #1.30
FILING DATE: 03-Apr-2001
CLASSIFICATION: US/09/825,574
PRIOR APPLICATION DATA: <Unknown>
APPLICATION NUMBER: 08/934,097
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION/DOCKET NUMBER: 38,230
TELECOMMUNICATION INFORMATION: FONS-02980
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
SEQUENCE FOR SEQ ID NO: 37:
LENGTH: 232
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: other nucleic acid
TOPOLOGY: linear
SEQUENCE DESCRIPTION: /desc = "DNA"

US-09-825-574-37/c
Sequence 37, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamichiev, Victor I.
Brow, Mary Ann D.
Neil, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
CORRESPONDENCE ADDRESS: 51
ADDRESS: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER: IBM PC compatible
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: Patentin Release #1.0, Version #1.30
FILING DATE: 03-Apr-2001
CLASSIFICATION: US/09/825,574
PRIOR APPLICATION DATA: <Unknown>
APPLICATION NUMBER: 08/934,097
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION/DOCKET NUMBER: 38,230
TELECOMMUNICATION INFORMATION: FONS-02980
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
SEQUENCE FOR SEQ ID NO: 37:
LENGTH: 232
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: other nucleic acid
TOPOLOGY: linear
SEQUENCE DESCRIPTION: /desc = "DNA"

US-09-825-574-37/c
Sequence 37, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamichiev, Victor I.
Brow, Mary Ann D.
Neil, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
CORRESPONDENCE ADDRESS: 51
ADDRESS: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER: IBM PC compatible
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: Patentin Release #1.0, Version #1.30
FILING DATE: 03-Apr-2001
CLASSIFICATION: US/09/825,574
PRIOR APPLICATION DATA: <Unknown>
APPLICATION NUMBER: 08/934,097
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION/DOCKET NUMBER: 38,230
TELECOMMUNICATION INFORMATION: FONS-02980
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
SEQUENCE FOR SEQ ID NO: 37:
LENGTH: 232
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: other nucleic acid
TOPOLOGY: linear
SEQUENCE DESCRIPTION: /desc = "DNA"

US-09-825-574-37/c
Sequence 37, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamichiev, Victor I.
Brow, Mary Ann D.
Neil, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
CORRESPONDENCE ADDRESS: 51
ADDRESS: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER: IBM PC compatible
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: Patentin Release #1.0, Version #1.30
FILING DATE: 03-Apr-2001
CLASSIFICATION: US/09/825,574
PRIOR APPLICATION DATA: <Unknown>
APPLICATION NUMBER: 08/934,097
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION/DOCKET NUMBER: 38,230
TELECOMMUNICATION INFORMATION: FONS-02980
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
SEQUENCE FOR SEQ ID NO: 37:
LENGTH: 232
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: other nucleic acid
TOPOLOGY: linear
SEQUENCE DESCRIPTION: /desc = "DNA"

US-09-825-574-37/c
Sequence 37, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamichiev, Victor I.
Brow, Mary Ann D.
Neil, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
CORRESPONDENCE ADDRESS: 51
ADDRESS: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER: IBM PC compatible
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: Patentin Release #1.0, Version #1.30
FILING DATE: 03-Apr-2001
CLASSIFICATION: US/09/825,574
PRIOR APPLICATION DATA: <Unknown>
APPLICATION NUMBER: 08/934,097
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION/DOCKET NUMBER: 38,230
TELECOMMUNICATION INFORMATION: FONS-02980
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
SEQUENCE FOR SEQ ID NO: 37:
LENGTH: 232
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: other nucleic acid
TOPOLOGY: linear
SEQUENCE DESCRIPTION: /desc = "DNA"

US-09-825-574-37/c
Sequence 37, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamichiev, Victor I.
Brow, Mary Ann D.
Neil, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
CORRESPONDENCE ADDRESS: 51
ADDRESS: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER: IBM PC compatible
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: Patentin Release #1.0, Version #1.30
FILING DATE: 03-Apr-2001
CLASSIFICATION: US/09/825,574
PRIOR APPLICATION DATA: <Unknown>
APPLICATION NUMBER: 08/934,097
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION/DOCKET NUMBER: 38,230
TELECOMMUNICATION INFORMATION: FONS-02980
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
SEQUENCE FOR SEQ ID NO: 37:
LENGTH: 232
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: other nucleic acid
TOPOLOGY: linear
SEQUENCE DESCRIPTION: /desc = "DNA"

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.

NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 239 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-825-574-36

Query Match 100.0%; Score 28; DB 10; Length 239;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCAAGCACCCCTATCAGCAGTACACAA 28
Db 239 GCAAGCACCCCTATCAGCAGTACACAA 212

RESULT 6
US-09-825-574-33/c
Sequence 33, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
Brow, Mary Ann D.
Fors, Lance
Neri, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 33:
US-09-825-574-33

Query Match 100.0%; Score 28; DB 10; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCAAGCACCCCTATCAGCAGTACACAA 28
Db 240 GCAAGCACCCCTATCAGCAGTACACAA 213

RESULT 7
US-09-825-574-34/c
Sequence 34, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
Brow, Mary Ann D.
Fors, Lance
Neri, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:

LENGTH: 240 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-825-574-34

Query Match 100.0%; Score 28; DB 10; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGGAGTACCAAA 28
|||||
DB 240 GCAAGCACCCTATCAGGAGTACCAAA 213

RESULT 8

US-09-825-574-35/C
Sequence 35, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:

APPLICANT: Lyamichay, Victor I.

Brow, Mary Ann D.
Fors, Lance
Neri, Bruce P.

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MERLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/825,574

FILING DATE: 03-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/934,097

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: MacKnight, Kamrin T.

REGISTRATION NUMBER: 38,230

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 35:

SEQUENCE CHARACTERISTICS:

LENGTH: 240 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-09-825-574-35

Query Match

100.0%; Score 28; DB 10; Length 240;

Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGGAGTACCAAA 28
|||||
DB 240 GCAAGCACCCTATCAGGAGTACCAAA 213

RESULT 9

US-10-087-631B-10/C
Sequence 10, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:

APPLICANT: JAEGER, STEPHAN

TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A

FILE REFERENCE: 1803-335-999

CURRENT APPLICATION NUMBER: US/10/087,631B

CURRENT FILING DATE: 2002-03-01

NUMBER OF SEQ ID NOS: 17

SOFTWARE: Patentln version 3.1

SEQ ID NO 10

LENGTH: 241

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: DNA sequence derived by

US-10-087-631B-10

Query Match 100.0%; Score 28; DB 9; Length 241;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGGAGTACCAAA 28
|||||
DB 241 GCAAGCACCCTATCAGGAGTACCAAA 214

RESULT 10

US-10-087-631B-13/C
Sequence 13, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:

APPLICANT: JAEGER, STEPHAN

TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A

FILE REFERENCE: 1803-335-999

CURRENT APPLICATION NUMBER: US/10/087,631B

CURRENT FILING DATE: 2002-03-01

NUMBER OF SEQ ID NOS: 17

SOFTWARE: Patentln version 3.1

SEQ ID NO 13

LENGTH: 241

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of artificial sequence: Amplicon sequence derived

OTHER INFORMATION: QS HCV (HCV amplification control having binding sites for ST2

US-10-087-631B-13

Query Match 100.0%; Score 28; DB 9; Length 241;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGGAGTACCAAA 28
|||||
DB 241 GCAAGCACCCTATCAGGAGTACCAAA 214

RESULT 11

US-10-087-631B-15/C

Sequence 15, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin version 3.1
SEQ ID NO 15
LENGTH: 242
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of artificial sequence: Amplicon derived from ICS62
OTHER INFORMATION: HCV (HCV-specific amplification control) using ST280 and ST778 as
US-10-087-631B-15

Query Match 100.0%; Score 28; DB 9; Length 242;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGCGAGTACCA 28
DB 242 GCAAGCACCCTATCAGCGAGTACCA 215

RESULT 12
US-09-825-574-26/c
Sequence 26, Application US/09825574
Patent No. US20020119454A1

GENERAL INFORMATION:

APPLICANT: Iyamichev, Victor I.
Brow, Mary Ann D.
Fors, Lance
Neri, Bruce P.

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/825,574

FILING DATE: 03-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/934,097

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: MacKnight, Kamrin T.

REGISTRATION NUMBER: 38,230

REFERENCE/DOCKET NUMBER: FOS-02980

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 244 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-09-825-574-26

Query Match 100.0%; Score 28; DB 10; Length 244;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGCGAGTACCA 28
DB 241 GCAAGCACCCTATCAGCGAGTACCA 214

RESULT 13

US-09-825-574-27/c
Sequence 27, Application US/09825574
Patent No. US20020119454A1

GENERAL INFORMATION:

APPLICANT: Iyamichev, Victor I.
Brow, Mary Ann D.
Fors, Lance
Neri, Bruce P.

TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides.

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/825,574

FILING DATE: 03-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/934,097

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: MacKnight, Kamrin T.

REGISTRATION NUMBER: 38,230

REFERENCE/DOCKET NUMBER: FOS-02980

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 244 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

SEQUENCE DESCRIPTION: SEQ ID NO: 27:
US-09-825-574-27

Query Match 100.0%; Score 28; DB 10; Length 244;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAGCACCCTATCAGCGAGTACCA 28
DB 241 GCAAGCACCCTATCAGCGAGTACCA 214

RESULT 14
US-09-825-574-28/c
Sequence 28, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
Brow, Mary Ann D.
Fors, Lance
Neri, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-09-825-574-28
Query Match 100.0%; Score 28; DB: 10; Length 244;
Best Local Similarity 100.0%; Pred. No. 0.001; 0; Indels 0; Gaps 0;
Matches 28; Conservative 0; Mismatches 0;
OY 1 GCAAGCACCTATCAGCAGTACCACAA 28
DB 241 GCAAGCACCTATCAGCAGTACCACAA 214
RESULT 15
US-09-825-574-29/c
Sequence 29, Application US/09825574
Patent No. US20020119454A1
GENERAL INFORMATION:
APPLICANT: Lyamichev, Victor I.
Brow, Mary Ann D.
Fors, Lance
Neri, Bruce P.
TITLE OF INVENTION: Polymorphism Analysis By Nucleic Acid
Structure Probing With Structure-Bridging
Oligonucleotides
NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/825,574
FILING DATE: 03-Apr-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/934,097
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: FORS-02980
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 29:
US-09-825-574-29
Query Match 100.0%; Score 28; DB: 10; Length 244;
Best Local Similarity 100.0%; Pred. No. 0.001; 0; Indels 0; Gaps 0;
Matches 28; Conservative 0; Mismatches 0;
OY 1 GCAAGCACCTATCAGCAGTACCACAA 28
DB 241 GCAAGCACCTATCAGCAGTACCACAA 214
Search completed: July 11, 2003, 15:01:57
Job time : 77.6382 secs

FEATURES	source
LOCUS	AL633783
DEFINITION	AL633783 XGC-gastrula <i>Silurana tropicalis</i> cDNA clone Tgas018013 5',
ACCESSION	AL633783
VERSION	AL633783.1
KEYWORDS	GI:16603266
SOURCE	EST.
ORGANISM	western clawed frog.
REFERENCE	<i>Silurana tropicalis</i>
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
JOURNAL	Xenopodinae; <i>Silurana</i> .
COMMENT	1 (bases 1 to 572)
	Huckle, E., Taylor, R., Ashurst, J. L., Zorn, A. M. and Rogers, J.
	Sanger Xenopus tropicalis EST project 2001 (10-2001)
	Unpublished (2001)
	Contact: Huckle E
	Sanger Centre
	Hinxton, Cambridgeshire, CB10 1SA, UK
	Email: tropesanger.ac.uk
	Sanger Xenopus tropicalis EST project 2001
	TROPICALIS_SEQUENCE_ID: Tgas018013.sp6
	Sequencing primer: SP6
	This sequence is from a Xenopus Gene Collection (XGC) library
	constructed by Aaron M. Zorn.
	Location/Qualifiers
	1..572
	/organism="Silurana tropicalis"

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 85.7%;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCAGGACCCCTACGACGACGACCA 28
Db 242 GCAGGACCCCTACGACGACGACCA 269

RESULT 2
AL637026 585 bp mRNA linear EST 12-DEC-2001
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Al637026
Al637026 XGC-neurula: Silurana tropicalis cDNA clone TNeu04n20 5',
mRNA sequence.
Al637026.1 GI:16789005
EST.
Western clawed frog.
Silurana tropicalis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodidae; Silurana.
1 (bases 1 to 585)
Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.
Unpublished (2001)
Contact: Huckle E
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE.ID: TNeu04n20.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1..585
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="TNeu04n20"
/clone_lib="XGC-neurula"
/dev_stage="neurula"
/lab_host="Escherichia coli DH10B"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 85.7%;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCAGGACCCCTACGACGACGACCA 28
Db 32 GCAGGACCCCTACGACGACGACCA 59

RESULT 3

AL793578 619 bp mRNA linear EST 27-JUN-2002
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Al793578
Al793578 XGC-neurula Silurana tropicalis cDNA clone TNeu11121 5',
mRNA sequence.
Al793578.1 GI:21579282
EST.
Western clawed frog.
Silurana tropicalis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodidae; Silurana.
1 (bases 1 to 619)
Taylor, R., Ashurst, J.L., Croning, M.D.R., Zorn, A.M. and Rogers, J.
Unpublished (2001)
Contact: Taylor R
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE.ID: TNeu11121.p1cSP6
Sequencing primer: p1cSP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1..619
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="TNeu11121"
/clone_lib="XGC-neurula"
/dev_stage="neurula"
/lab_host="Escherichia coli DH10B"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 85.7%;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCAGGACCCCTACGACGACGACCA 28
Db 61 GCAGGACCCCTACGACGACGACCA 88

RESULT 4
B177787/c 306 bp mRNA linear EST 23-JUL-2002
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

B177787
B177787 EBR08_SQ002_B06.R root, 3 week, drought-stressed, cv Optic, EBR08
Hordeum vulgare cDNA clone EBR08_SQ002_B06 5', mRNA sequence.
B177787.2 GI:21951028
EST.
Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
1 (bases 1 to 306)
Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.
Development of Barley Transcriptome Resources
Unpublished (2001)
On Sep 26, 2001 this sequence version replaced gi:15780679.
Contact: Waugh R, Marshall DR
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731

Query Match 75.7% Score 21.2; DB 14; Length 731;
Best Local Similarity 88.5% Pred. No. 89;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCAGCACCCTATCAGCAGTACCAC 26
DB 381 GCAGCTCCGACAGCAGTACCAC 356
|||||

RESULT 7
BHI01607/c
LOCUS
DEFINITION BHI01607 482 bp DNA linear GSS 19-JUN-2001
RPCI-24-229K3.TJ RPCI-24 Mus musculus genomic clone RPCI-24-229K3.
ACCESSION
VERSION BHI01607
KEYWORDS BHI01607.1 GI:14926660
SOURCE GSS.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akintet, B., Levins, M.,
Tsegaye, G., Geer, K., Kroll, M., Shvartsbeyn, A., Gebregeorgis, E.,
Russell, D., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other GSSs: RPCI-24-229K3.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaot@ig.org

TITLE
JOURNAL
COMMENT
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@email.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end plate: http://www.choi.org/tb/bac_ends/mouse/bac_end_intro.html
Seq primer: SP6
Class: BAC ends.

Location/Qualifiers
1.482
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-229K3"
/clone.lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTRABAC1, Site 1: BamHI, Site 2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."
138 a 87 c 99 g 158 t

FEATURES
Source

BASE COUNT
ORIGIN

Query Match 73.6% Score 20.6; DB 17; Length 482;
Best Local Similarity 85.2% Pred. No. 1,4e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 2 CAAGCACCCTATCAGCAGTACCACAA 28
DB 462 CAAGCACCCTATCAGCAGTACCACAA 436
|||||

RESULT 8
LOCUS
AI657832
560 bp mRNA linear EST 07-JUN-2001

DEFINITION

fc23e12.y1 zebrafish Washu MPing EST Danio rerio cDNA clone
IMAGE:3722254.5 similar to SW-1YHF SCHPO 044360 HYPOTHETICAL 57.6
KD PROTEIN C30D10.15 IN CHROMOSOME II.1, mRNA sequence.
AI657832
AI657832.1 GI:4755500
EST.

KEYWORDS

zebrafish.

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
Cyprinidae; Danio.
1 (bases 1 to 560)

REFERENCE

AUTHORS
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wyle, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Rittler, E., Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.
Washu zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zdrafish@wustl.edu

TITLE

JOURNAL

COMMENT

CDNA Library Preparation: Matthew Clark. CDNA Library Arrayed by: Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and ResourcenZentrumPrimarDatenbank, Berlin, Germany (web address: www.rzpd.de)
Seq primer: T3 ET from Amersham
High quality sequence stop: 492.

FEATURES
Source

Location/Qualifiers
1.560
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="IMAGE:3722254"
/clone.lib="zebrafish Washu MPing EST"
/sex="mixed"
/tissue.type="26 somite embryos, adult livers, shield stage embryos"
/lab.host="XLI-blue MRF"
/note="Vector: pSPORT1, Site 1: NotI, Site 2: SalI; 1st strand cDNA was primed with a Not I - oligo(dT)15 primer (5'-GACAGTCTTCTAGATCGGAGCGCCCTTTTCTTTTCTTTT3'); double-stranded cDNA was ligated to Sal I adaptors (BRU), digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRU). Library was constructed by Matthew Clark (Lehrach lab, ICRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

BASE COUNT
ORIGIN
162 a 163 c 122 g 113 t

Query Match 71.4% Score 20; DB 9; Length 560;
Best Local Similarity 82.1% Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GCAGCACCCTATCAGCAGTACCACAA 28
|||||

Db 417 GCAACACACCTCGTGGCAGTACCACAA 444

LOCUS	RESULT 9	LOCUS
EG083128		
DEFINITION	H3084G10-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone	
ACCESSION	H3084G10 5', mRNA sequence.	
VERSION	EG083128	
KEYWORDS	EG083128.1 GI:12565696	
SOURCE	EST.	
ORGANISM	house mouse.	
REFERENCE	Mus musculus	
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 595)	
JOURNAL	Kargul, G. J., Dudekula, D. B., Qian, Y., Lam, M. K., Jaradat, S. A., Tanaka, R. S., Carter, M. G. and Ko, M. S. H.	
COMMENT	Verification and initial annotation of NIA mouse 15K cDNA clone set Unpublished (2001)	
	Other-ESTs: H3084G10-3	
	Contact: George J. Kargul	
	Laboratory of Genetics	
	National Institute on Aging/National Institutes of Health	
	333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA	
	Email: cdna@igsun.grc.nia.nih.gov	
	This clone set has been freely distributed to the community. Please visit http://igsun.grc.nia.nih.gov/CDNA/15k.html for details.	
	Plate: H3084 row: G column: 10	
	Seq primer: -21M13 Reverse	
	High quality sequence stop: 595	
	POLYA-No.	

```

FEATURES
SOURCE
Location/Qualifiers
1..595
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taeST:H3084G10-5"
/db_xref="taxon:10090"
/clone="H3084G10"
/clone_11b="NIA Mouse 15K CDNA Clone Set"
/sex="Clones arrayed from a variety of CDNA libraries"
/idx_stage="Clones arrayed from a variety of CDNA libraries"
/lab_host="DH10B"
/notes="Vector: pSPORT1; Site.1: SalI; Site.2: NotI. This clone is among a rearranged set of 15,247 clones from 11 embryo CDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary CDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dTm) -Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."
Mol Genet 7: 1967-1978."
BASE COUNT 143 a 177 c 112 g 163 t
ORIGIN
Query Match 71.4% Score 20; DB 12; Length 595;
Best Local Similarity 82.1% Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
1 GCAGCACCCTATCAGCAGATCACCA 28
11 ||||||| ||||| |||||
501 GCTGACACCTCTTAGCGCATCCACA 578

```

[illegible][illegible]

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA.
Email: cdna@gsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit <http://lgsun.grc.nia.nih.gov/cdna/15k.html> for details.
Plate: H3084 row: 6 column: 10
Seq primer: -21M13 Forward
High quality sequence stop: 827
POLA-yes

FEATURES

SOURCE

Location/Qualifiers
1. 827
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:H3084G10-3"
/db_xref="taxon:10090"
/clone="H3084G10"
/clone_lib="NIA Mouse 15k cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"

BASE COUNT

211 a 184 c 203 g 229 t

ORIGIN

Query Match

Best Local Similarity 82.1%; Score 20; DB 12; Length 827;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

DB 1 GCAAGACCCCTATCAGCAGTACACAA 28
425 GCTAGCACCCTCTAAGCAGATCCACAA 398

RESULT 12

LOCUS BG953300 235 bp mRNA linear EST 12-JUN-2001
DEFINITION CM4-CT0629-220101-700-909 CT0629 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG953300
VERSION BG953300.1 GI:14371471
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 235)

REFERENCE

1 (bases 1 to 235)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Negal, M.A., da Silva, W. Jr., Zago, M.A., Borlin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

JOURNAL Proc Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the RAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4&t2=CM4-CT0629-220101-700-909&t3=2001-01-22&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 124
High quality sequence stop: 187.

FEATURES

SOURCE

Location/Qualifiers
1. 235
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0629"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site: 1; SmaI; Site: 2;
SmaI: A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT

77 a 68 c 47 g 43 t

ORIGIN

Query Match
Best Local Similarity 84.6%; Score 19.6; DB 13; Length 235;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

DB 3 AAGCAGCTATCTGCGAGTACACAA 28
41 AAGCAGCTATCTGCGAGTACACAA 66

RESULT 13

LOCUS BE558358 828 bp mRNA linear EST 23-OCT-2001
DEFINITION HV.CEBD017F16f Hordeum vulgare seedling green leaf EST library
ACCESSION HVCDBA0005 (Blumeria challenged) Hordeum vulgare cDNA clone
VERSION BE558358
KEYWORDS EST.
SOURCE BE558358.2 GI:13265876
ORGANISM Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
1 (bases 1 to 828)

REFERENCE

1 (bases 1 to 828)
Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Wei, F., Begum, D.,
Fritsch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi,
D.W., Fenton, R.D., Oates, R. and Main, D.
Development of a genetically and physically anchored EST resource
for barley genomics: Blumeria infected incompatible (Mia6) seedling
leaf cDNA library
Unpublished (2001)

TITLE

JOURNAL

On Aug 14, 2000 this sequence version replaced gi:9822848.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 113
Seq primer: AATTAACCTCCTAAGCG
High quality sequence start: 4

Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mousestewatson.wustl.edu

This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (infoimage.lnl.gov) for further information.
 MGI:969913

Seq primer: custom primer used

FEATURES

High quality sequence stop: 450.
 Location/Qualifiers

```

1..475
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1885589"
/dev_stage="embryo, 14 dpc"
/lab_host="DH10B"
/note="Vector: pME18S-FL3; Site 1: DraIII (CACGTGTG);
Site 2: DraIII (CACCATGTG); 1st strand cDNA was primed
with an oligo(dT) primer (ATGTGGCTTTTCTTTTCTTTT);
double-stranded cDNA was ligated to a DraIII adaptor
[GTGTGGCTTACTGCG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
insert). Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano.
(University of Tokyo, Institute of Medical Science).
Custom primers for sequencing: 5' end primer
CTTCTGCTCTAAAGTCGCG and 3' end primer
CGACCTGAGCTGCACACA."
BASE COUNT      149 a      80 c      91 g      154 t      1 others
ORIGIN

```

Query Match
 Best Local Similarity 81.9% Score 19; DB 9; Length 475;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

OY      1  GCAGCAGCCTATCAGGAGTACCA 27
          |||||
Db      341 GCAGCAGCATGACAGCAGACCA 367

```

Search completed: July 11, 2003 02:25:58
 Job time : 511.434 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 225.294 Seconds

(without alignments)
3358.597 Million cell updates/sec

Title: US-10-087-631b-8

Perfect score: 26

Sequence: 1 cgtcttcgcagatcgctacat 26

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database :

Listing first 45 summaries

GenEmbl1:
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_pl:*
8: gb_pl:*
9: gb_pl:*
10: gb_pl:*
11: gb_pl:*
12: gb_pl:*
13: gb_pl:*
14: gb_pl:*
15: gb_pl:*
16: gb_pl:*
17: gb_pl:*
18: gb_pl:*
19: gb_pl:*
20: gb_pl:*
21: gb_pl:*
22: gb_pl:*
23: gb_pl:*
24: gb_pl:*
25: gb_pl:*
26: gb_pl:*
27: gb_pl:*
28: gb_pl:*
29: gb_pl:*
30: gb_pl:*
31: gb_pl:*
32: gb_pl:*
33: gb_pl:*
34: gb_pl:*
35: gb_pl:*
36: gb_pl:*
37: gb_pl:*
38: gb_pl:*
39: gb_pl:*
40: gb_pl:*
41: gb_pl:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.2	73.8	139534	2	AP005640
2	18.6	71.5	106284	2	AC122161
3	18.6	71.5	112538	2	AC122171
4	18.6	71.5	145424	2	AC114235
5	18.6	71.5	198868	2	AC110415
6	18.6	71.5	208396	2	AC003101
7	18.6	71.5	212030	2	AC019148
8	17.8	68.5	1290	3	AF358892
9	17.8	68.5	1290	3	AF358893
10	17.8	68.5	1290	3	AF358894
11	17.8	68.5	1290	3	AF358895
12	17.8	68.5	1290	3	AF358897
13	17.8	68.5	1290	3	AF358899
14	17.8	68.5	1290	3	AF359001
15	17.8	68.5	1290	3	AF359003
16	17.8	68.5	1290	3	AF359005
17	17.8	68.5	1290	3	AF359005
18	17.8	68.5	13051	1	AE009111
19	17.8	68.5	14856	1	AE008077
20	17.8	68.5	20389	1	AE008768
21	17.8	68.5	70100	8	AP001306
22	17.8	68.5	315000	8	RME603644
23	17.6	67.7	1333	8	AY128313
24	17.6	67.7	2191	8	AY039862
25	17.6	67.7	2638	8	AY051704
26	17.6	67.7	3273	8	PM000000
27	17.6	67.7	39780	2	AC014011
28	17.6	67.7	54115	2	AC018149
29	17.6	67.7	60040	3	AC005128
30	17.6	67.7	76170	25	ATAC3329
31	17.6	67.7	85660	9	AC092845
32	17.6	67.7	92085	2	AC130938
33	17.6	67.7	95613	2	AC020389
34	17.6	67.7	99725	8	ATF18P4
35	17.6	67.7	101647	8	AC004665
36	17.6	67.7	110726	2	AC131225
37	17.6	67.7	111705	2	AC096275
38	17.6	67.7	117080	2	AC098840
39	17.6	67.7	128903	2	AC093952
40	17.6	67.7	139367	2	AC105676
41	17.6	67.7	151610	3	AC009462
42	17.6	67.7	152021	2	AC010923
43	17.6	67.7	153480	2	AC130879
44	17.6	67.7	155557	3	AC099020
45	17.6	67.7	155840	2	AC006495

ALIGNMENTS

RESULT 1
AP005640
LOCUS
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 2 clone
OSJNB0006015. *** SEQUENCING IN PROGRESS ***. In ordered pieces.
ACCESSION AP005640.1 GI:22266684
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Baylor Plaza, Houston, TX 77030, USA
(bases 1 to 145424)

Submitted (13-JUL-2002)

Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:19224562.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project information

Center project name: GQW

Center clone name: CH230-334N2

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 118496 bases at least Q40

Consensus quality: 121465 bases at least Q30

Consensus quality: 123679 bases at least Q20

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

NOTE: This is a 'working draft' sequence. It currently

consists of 40 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1181: contig of 1181 bp in length
1182: contig of 1182 bp in length
1282: contig of 1247 bp in length
2529: gap of unknown length
2529: gap of unknown length
3792: contig of 1163 bp in length
3892: gap of unknown length
5365: contig of 1473 bp in length
5465: gap of unknown length
6664: contig of 1200 bp in length
6764: gap of unknown length
8080: contig of 1316 bp in length
8180: gap of unknown length
8181: contig of 1069 bp in length
9249: gap of unknown length
9350: contig of 1162 bp in length
10511: gap of unknown length
10611: contig of 1721 bp in length
12332: gap of unknown length
12432: gap of unknown length
14518: contig of 2086 bp in length
14618: gap of unknown length
14619: contig of 1834 bp in length
16552: gap of unknown length
18483: contig of 1931 bp in length
18584: gap of unknown length
21569: contig of 2986 bp in length
21669: gap of unknown length
24031: contig of 2362 bp in length
24131: gap of unknown length
26242: contig of 2111 bp in length
26342: gap of unknown length
27981: contig of 1639 bp in length
28081: gap of unknown length
30840: contig of 2759 bp in length
30940: gap of unknown length
30941: contig of 2888 bp in length
33829: gap of unknown length
33928: contig of 1599 bp in length
3527: gap of unknown length
35628: contig of 2930 bp in length

FEATURES

source

38558 38657: gap of unknown length
38658 41510: contig of 2853 bp in length
41511 41610: gap of unknown length
41611 44279: contig of 2569 bp in length
44280 44379: gap of unknown length
44380 48860: contig of 4481 bp in length
48861 53123: contig of 4163 bp in length
53124 53224: gap of unknown length
53224 56643: contig of 3420 bp in length
56644 56743: gap of unknown length
56744 60605: contig of 3862 bp in length
60606 60705: gap of unknown length
60706 64174: contig of 3469 bp in length
64175 64274: gap of unknown length
64275 68279: contig of 4005 bp in length
68280 73156: contig of 4777 bp in length
73157 73256: gap of unknown length
73257 77125: contig of 3869 bp in length
77126 77225: gap of unknown length
77226 81324: contig of 3999 bp in length
81325 85376: gap of unknown length
85377 85676: contig of 4252 bp in length
85677 89576: gap of unknown length
89578 95097: contig of 4081 bp in length
89579 95098: gap of unknown length
95099 95197: gap of unknown length
95198 100912: contig of 5240 bp in length
100913 101012: contig of 5715 bp in length
101013 107906: gap of unknown length
107907 108006: contig of 6894 bp in length
108007 114369: gap of unknown length
114370 114469: gap of unknown length
114471 121563: contig of 7094 bp in length
121564 121663: gap of unknown length
121664 134667: contig of 13004 bp in length
134668 134767: gap of unknown length
134768 145424: contig of 10657 bp in length.

Location/Qualifiers

1. 145424
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-334N2"

BASE COUNT 38722 a 30549 c 31478 g 39457 t 5218 others

Query Match Best Local Similarity 71.58; Score 18.6; DB 2; Length 145424;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OR 2 GTCCTTCGACGATCGTCAAT 26

DB 89203 GTCCTTCGACGATCGTCAAT 89179

RESULT 5 AC110415 198868 bp. DNA linear HTG 17-JUL-2002

LOCUS AC110415 Rattus norvegicus clone CH230-126C16, *** SEQUENCING IN PROGRESS

DEFINITION *** 58 unordered pieces.

AC110415 AC110415.3 GT:21745530

VERSION HTG; HTGS-PHASE1.

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 198868)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C.,

Alstirooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,

Barbata, J., Benton, J., Blincke, K., Blankenburg, K., Bonini, D.,
 Bouck, J., Boyle, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
 Burch, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.,
 Donath, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Eathart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gortelli, J.H., Guevara, N., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homsi, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, T., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratoch, J., Kueshl, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenwo, S., Ogun, M., Okunnu, G.,
 Orquenie, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,
 Sodergren, E., Soneike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, Y., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, J., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 198868)
 Worley, K.C.
 Direct Submission
 Submitted (12-FEB-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 198868)
 Worley, K.C.
 Direct Submission
 Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 14, 2002 this sequence version replaced gi:18851028.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GSRL
 Center clone name: CH230-126C16
 ----- Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 149839 bases at least Q40
 Consensus quality: 156891 bases at least Q30
 Consensus quality: 162963 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 58 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	1081: contig of 1081 bp in length
1082	1181: gap of unknown length
1182	2375: contig of 1194 bp in length
2376	2475: gap of unknown length
2476	3595: contig of 1120 bp in length
3596	3695: gap of unknown length
3696	4903: contig of 1208 bp in length
4904	5003: gap of unknown length
5004	6036: contig of 1033 bp in length
6037	6136: gap of unknown length
6137	7512: contig of 1376 bp in length
7513	7612: gap of unknown length
7613	8924: contig of 1312 bp in length
8925	9024: gap of unknown length
9025	10149: contig of 1125 bp in length
10150	10249: gap of unknown length
10250	11272: contig of 1023 bp in length
11273	11372: gap of unknown length
11373	13108: contig of 1736 bp in length
13109	13208: gap of unknown length
13209	15193: contig of 1985 bp in length
15194	15293: gap of unknown length
15294	16884: contig of 1551 bp in length
16885	16984: gap of unknown length
16985	19379: contig of 2255 bp in length
19380	19379: gap of unknown length
19380	20807: contig of 1428 bp in length
20808	20907: gap of unknown length
20908	23046: contig of 2139 bp in length
23047	23146: gap of unknown length
23147	24865: contig of 1719 bp in length
24866	24965: gap of unknown length
24966	26813: contig of 1848 bp in length
26814	26913: gap of unknown length
26914	29558: contig of 2745 bp in length
29559	29758: gap of unknown length
29759	31694: contig of 1936 bp in length
31695	31794: gap of unknown length
31795	33986: contig of 2192 bp in length
33987	34086: gap of unknown length
34087	36029: contig of 1943 bp in length
36030	36129: gap of unknown length
36130	38429: contig of 2300 bp in length
38430	38529: gap of unknown length
38530	40701: contig of 2172 bp in length
40702	40801: gap of unknown length
40802	42597: contig of 1796 bp in length
42598	42697: gap of unknown length
42698	44891: contig of 2194 bp in length
44892	44991: gap of unknown length
44992	48321: contig of 3330 bp in length
48322	48421: gap of unknown length
48422	50838: contig of 2417 bp in length
50839	50938: gap of unknown length
50939	52492: contig of 1554 bp in length
52493	52592: gap of unknown length
52593	55492: contig of 2900 bp in length
55493	55592: gap of unknown length
55593	57986: contig of 2394 bp in length
57987	58086: gap of unknown length
58087	60471: contig of 2385 bp in length
60472	60571: gap of unknown length
60572	64863: contig of 4292 bp in length
64864	64963: gap of unknown length
64964	67939: contig of 2976 bp in length
67940	68039: gap of unknown length
68040	71114: contig of 3075 bp in length


```

repeat_region complement(10288. .10591)
/rpt_family="AluSb"
repeat_region complement(10592. .10820)
/rpt_family="L1M4"
repeat_region 10861. .11163
/rpt_family="AluSp"
repeat_region complement(11168. .11260)
/rpt_family="L1M4"
repeat_region complement(11280. .11573)
/rpt_family="AluSx"
repeat_region complement(11615. .11917)
/rpt_family="AluSg"
repeat_region 11957. .12251
/rpt_family="AluSb"
repeat_region complement(12253. .12681)
/rpt_family="L1M4"
repeat_region complement(12701. .12956)
/rpt_family="L1M4"
repeat_region complement(12976. .13276)
/rpt_family="AluSg"
repeat_region complement(13277. .13417)
/rpt_family="L1M4"
repeat_region complement(13424. .13606)
/rpt_family="AluSb"
repeat_region complement(13657. .13911)
/rpt_family="L1M4"
repeat_region complement(14096. .14403)
/rpt_family="L1M4"
repeat_region complement(14527. .14634)
/rpt_family="AluSg"
repeat_region complement(14635. .14933)
/rpt_family="AluSx"
repeat_region 15106. .15347
/rpt_family="MIR"
repeat_region complement(16322. .16354)
/rpt_family="AluSg"
repeat_region 16374. .16492
/rpt_family="MIR"
repeat_region complement(16557. .16957)
/rpt_family="AluSg"
repeat_region complement(17013. .17317)
/rpt_family="AluSx"
repeat_region 18703. .18766
/rpt_family="L1"
repeat_region 18855. .19009
/rpt_family="FRAM"
repeat_region complement(19035. .19146)
/rpt_family="AluSg"
repeat_region complement(19177. .19309)
/rpt_family="L1"
repeat_region 19603. .19906
/rpt_family="AluSx"
repeat_region complement(20121. .20414)
/rpt_family="L1"
repeat_region complement(20415. .20717)
/rpt_family="AluSg"
repeat_region complement(20719. .20808)
/rpt_family="L1"
repeat_region complement(21025. .21151)
/rpt_family="L1"
repeat_region 22253. .22356
/rpt_family="MER81"
repeat_region complement(22414. .22718)
/rpt_family="AluSb"
repeat_region 22901. .23218
/rpt_family="AluSb"
repeat_region 24532. .24942
/rpt_family="MER21B"
repeat_region 24963. .25259
/rpt_family="AluSb"
repeat_region 25269. .25384
/rpt_family="MER21B"
repeat_region 26217. .26321

```

```

repeat_region /rpt_family="MIR"
complement(26787. .26834)
repeat_region /rpt_family="(CAAA)n"
complement(26835. .27117)
repeat_region /rpt_family="AluY"
28880. .29005
/rpt_family="FRAM_C"
repeat_region complement(29274. .29658)
/rpt_family="L1M4"
repeat_region complement(29660. .29977)
/rpt_family="AluSx"
repeat_region complement(30027. .30526)
/rpt_family="L1M4"
repeat_region 30536. .30837
/rpt_family="AluSx"
repeat_region complement(30938. .31065)
/rpt_family="L1M4"
repeat_region complement(31068. .31374)
/rpt_family="AluSg"
repeat_region complement(31576. .31876)
/rpt_family="AluSg"
repeat_region complement(32053. .32347)
/rpt_family="AluY"
repeat_region complement(32460. .32677)
/rpt_family="MER58A"
repeat_region 32727. .33028
/rpt_family="AluSx"
repeat_region 33679. .33985
/rpt_family="AluSx"
repeat_region complement(35256. .35582)
/rpt_family="L1M4"
repeat_region complement(35602. .35906)
/rpt_family="AluSg"

```

```

Query Match 71.5%; Score 18.6; DB 9; Length 208396;
Best Local Similarity 84.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CGCTTCGACGATGCTACCTCA 25
DB 152067 CCTCTTCACAGATGCTACCTCA 152091

```

```

RESULT 7
AC019148 212030 bp DNA linear HTG 19-JUL-2000
LOCUS Homo sapiens chromosome 17 clone RP11-772K16, WORKING DRAFT
DEFINITION
SEQUENCE, 15 unordered pieces.
ACCESSION
AC019148
VERSION
AC019148.4 GI:9230872
KEYWORDS
HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa;
REFERENCE
1 (bases 1 to 212030)
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS
Waterston, R.H.
JOURNAL
The sequence of Homo sapiens clone
Unpublished
TITLE
2 (bases 1 to 212030)
REFERENCE
Waterston, R.H.
AUTHORS
Direct Submission
TITLE
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
JOURNAL
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Jul 16, 2000 this sequence version replaced gi:7940422.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0772K16

```

Summary Statistics

Sequencing vector: M13, 99%
 Chemistry: Dye-terminator Big Dye, 1% of reads
 Assembly program: Phrap, version 0.990319
 Consensus quality: 200508 bases at least Q40
 Consensus quality: 20488 bases at least Q30
 Insert size: 25800; agarose-gel
 Insert size: 210630; sum-of-contigs
 Quality coverage: 4.30 in Q20 bases; agarose-gel
 Quality coverage: 4.57 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1597 1596: contig of 1596 bp in length
 1697 1696: gap of unknown length
 3382 3381: contig of 1685 bp in length
 3482 3481: gap of unknown length
 5703 5702: contig of 2221 bp in length
 5803 5802: gap of unknown length
 7840 7839: contig of 2037 bp in length
 7940 7939: gap of unknown length
 11444 11443: contig of 3204 bp in length
 11244 11243: gap of unknown length
 14660 14659: contig of 3416 bp in length
 14760 14759: gap of unknown length
 19765 19764: contig of 5005 bp in length
 19865 19864: gap of unknown length
 28811 28810: contig of 8946 bp in length
 28911 28910: gap of unknown length
 42898 42897: contig of 13987 bp in length
 42998 42997: gap of unknown length
 59240 59239: contig of 16242 bp in length
 59340 59339: gap of unknown length
 73504 73503: contig of 14164 bp in length
 73604 73603: gap of unknown length
 89176 89175: contig of 15572 bp in length
 89276 89275: gap of unknown length
 118498 118497: contig of 29222 bp in length
 118598 118597: gap of unknown length
 156775 156774: contig of 38077 bp in length
 156775 212030: contig of 55256 bp in length

Location/Qualifiers
 1. 212030
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="17"
 /clone="RP11-772K16"
 1. 1596
 /note="assembly_name:Contig4"
 1697 3381
 /note="assembly_name:Contig5"
 3482 5702
 /note="assembly_name:Contig6"
 5803 7839
 /note="assembly_name:Contig7"
 7940 11143
 /note="assembly_name:Contig8"
 11244 14659
 /note="assembly_name:Contig9"
 14760 19764
 /note="assembly_name:Contig10"
 clone_end:sp6
 vector_side:left"

FEATURES

misc_feature
 1. 1596
 /note="assembly_name:Contig4"
 1697 3381
 /note="assembly_name:Contig5"
 3482 5702
 /note="assembly_name:Contig6"
 5803 7839
 /note="assembly_name:Contig7"
 7940 11143
 /note="assembly_name:Contig8"
 11244 14659
 /note="assembly_name:Contig9"
 14760 19764
 /note="assembly_name:Contig10"
 clone_end:sp6
 vector_side:left"

misc_feature 19865..28810
 /note="assembly_name:Contig11"
 misc_feature 28911..42897
 /note="assembly_name:Contig12"
 misc_feature 42998..59239
 /note="assembly_name:Contig13"
 misc_feature 59340..73503
 /note="assembly_name:Contig14"
 misc_feature 73604..89175
 /note="assembly_name:Contig15"
 misc_feature 89276..118497
 /note="assembly_name:Contig16"
 misc_feature 118598..156774
 /note="assembly_name:Contig17"
 misc_feature 156775..212030
 /note="assembly_name:Contig18"
 BASE COUNT 57448 a 48441 c 49876 g 54852 t 1413 others
 ORIGIN

Query Match 71.5% Score 18.6; DB 2; Length 212030;
 Best Local Similarity 84.0%; Pred. No. 1.3e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 CGCTTCGCGAGTCGCTACTCA 25
 Db 101558 CGCTTCGCGAGTCGCTACTCA 101582

RESULT 8
 LOCUS AF358992/c 1290 bp DNA linear INV 05-JUL-2001
 DEFINITION Trypanosoma cruzi strain ESMERALDO c13 trypanothione reductase (TR)
 ACCESSION AF358992
 VERSION AF358992.1 GI:14600021
 KEYWORDS
 SOURCE Trypanosoma cruzi.
 ORGANISM Trypanosoma cruzi.
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma; Schizotrypanum.

REFERENCE 1 (bases 1 to 1290)
 Trypanosoma; Schizotrypanum.
 MACHADO, C.A. and AYALA, F.J.
 Nucleotide sequences provide evidence of genetic exchange among
 distinct related lineages of Trypanosoma cruzi
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001)
 PUBMED 11416213
 MEDLINE 21310003
 JOURNAL 21310003
 AUTHORS Machado, C.A. and Ayala, F.J.
 TITLE Submitted (12-MAR-2001) Ecology and Evolutionary Biology,
 University of California, Irvine, 321 Steinhaus Hall, Irvine, CA
 JOURNAL 92697-2525, USA

FEATURES

source
 1. 1290
 /organism="Trypanosoma cruzi"
 /strain="ESMERALDO c13"
 /db_xref="taxon:5693"
 <1..>1290
 /gene="tr"
 <1..>1290
 /gene="trp"
 /product="trypanothione reductase"
 <1..>1290
 /gene="tr"
 /codon_start=1
 /product="trypanothione reductase"
 /protein_id="AAK71210.1"
 /db_xref="GI:14600022"
 /translation="TLTKRRAVAVDVGVHGPPEFSAIGSTGVNVCYKRLKMTGAO
 GSLESKNVNAESADPASVAKERLEKLLIAVDEAVIYNKSYEMERFDEGELEFLGW
 PEPFRVLTGGGIFSEFAIGFNAVKPGCVTLCTRGEMILIRGPHLREELTKOL"

TANGIOILITKENPAKVELINDGSKSVTEESGKKMDFLVMAIGRSPTKDLQONAG
VMKNGGVQVDEYSRTNYSNITAIAGDVTNRMLPVPVAINEAALVDYFETNPKRTOH
TRVSAVSFTIPGICGLIEEVAASKRREYVAVYSSFTPLMHNISGSKYKTFVAKIIT
NHSOGTVIGVHLIDGNAPEIIIGVGICIKLNAKISDFY"

BASE COUNT

309 a 318 c 353 g 310 t

Query Match 68.5%; Score 17.8; DB 3; Length 1290;
Best Local Similarity 90.5%; Pred. No. 3e+02; 2; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGCTTTGCGAGATCGTACC 21
Db 996 CGCTTTGCGAGATCGTACC 976

RESULT 9
AF358993/c 1290 bp DNA linear INV 05-JUL-2001

LOCUS Trypanosoma cruzi strain T018 cl2 trypanothione reductase (TR)
DEFINITION gene, partial cds.

ACCESSION AF358993.1 GI:14600023

VERSION AF358993.1

KEYWORDS

SOURCE

ORGANISM

Trypanosoma cruzi.
Trypanosoma cruzi.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.

REFERENCE
AUTHORS Machado, C.A. and Ayala, F.J.
TITLE Nucleotide sequences provide evidence of genetic exchange among
distantly related lineages of Trypanosoma cruzi
Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001)

JOURNAL 21310003
MEDLINE 11416213
PUBMED 11416213
REFERENCE 2 (bases 1 to 1290)
AUTHORS Machado, C.A. and Ayala, F.J.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2001) Ecology and Evolutionary Biology,
University of California, Irvine, 321 Steinhaus Hall, Irvine, CA
92697-2525, USA

FEATURES
Source Location/Qualifiers
1..1290
/organism="Trypanosoma cruzi"
/strain="T018 cl2"
/db_xref="taxon:5693"
<1..>1290
/gene="TR"
<1..>1290
/gene="TR"
/product="trypanothione reductase"
<1..>1290
/gene="TR"
/codon_start=1
/product="trypanothione reductase"
/protein_id="AAK71211.1"
/db_xref="GI:14600023"

gene
mRNA
CDS
BASE COUNT 309 a 318 c 353 g 310 t

ORIGIN
Query Match 68.5%; Score 17.8; DB 3; Length 1290;
Best Local Similarity 90.5%; Pred. No. 3e+02; 2; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGCTTTGCGAGATCGTACC 21

||||||| ||| |||||
Db 996 CGCTTTGCGAGATCGTACC 976

RESULT 10
AF358994/c 1290 bp DNA linear INV 05-JUL-2001

LOCUS Trypanosoma cruzi strain CBB cl3 trypanothione reductase (TR) gene,
partial cds.

ACCESSION AF358994.1 GI:14600025

VERSION AF358994.1

KEYWORDS

SOURCE

ORGANISM

Trypanosoma cruzi.
Trypanosoma cruzi.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.

REFERENCE
AUTHORS Machado, C.A. and Ayala, F.J.
TITLE Nucleotide sequences provide evidence of genetic exchange among
distantly related lineages of Trypanosoma cruzi
Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001).

JOURNAL 21310003
MEDLINE 11416213
PUBMED 11416213
REFERENCE 2 (bases 1 to 1290)
AUTHORS Machado, C.A. and Ayala, F.J.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2001) Ecology and Evolutionary Biology,
University of California, Irvine, 321 Steinhaus Hall, Irvine, CA
92697-2525, USA

FEATURES
Source Location/Qualifiers
1..1290
/organism="Trypanosoma cruzi"
/strain="CBB cl3"
/db_xref="taxon:5693"
<1..>1290
/gene="TR"
<1..>1290
/gene="TR"
/product="trypanothione reductase"
<1..>1290
/gene="TR"
/codon_start=1
/product="trypanothione reductase"
/protein_id="AAK71211.1"
/db_xref="GI:14600025"

gene
mRNA
CDS
BASE COUNT 309 a 318 c 353 g 310 t

ORIGIN
Query Match 68.5%; Score 17.8; DB 3; Length 1290;
Best Local Similarity 90.5%; Pred. No. 3e+02; 2; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGCTTTGCGAGATCGTACC 21
Db 996 CGCTTTGCGAGATCGTACC 976

RESULT 11
AF358995/c 1290 bp DNA linear INV 05-JUL-2001

LOCUS Trypanosoma cruzi strain MSC2 trypanothione reductase (TR) gene,
partial cds.

ACCESSION AF358995.1 GI:14600027

VERSION AF358995.1

KEYWORDS

REFERENCE	JOURNAL	REFERENCE	JOURNAL
AUTHORS	MEDLINE	AUTHORS	MEDLINE
TITLE	PUBMED	TITLE	PUBMED
ORGANISM	ORGANISM	ORGANISM	ORGANISM
1 (bases 1 to 1290)	21310003	1 (bases 1 to 1290)	21310003
Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of <i>Trypanosoma cruzi</i>	11416213	Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of <i>Trypanosoma cruzi</i>	11416213
Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001)		Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001)	
2 (bases 1 to 1290)		2 (bases 1 to 1290)	
Machado, C.A. and Ayala, F.J.		Machado, C.A. and Ayala, F.J.	
Direct Submission		Direct Submission	
Submitted (12-MAR-2001) Ecology and Evolutionary Biology, University of California, Irvine, 321 Steinhaus Hall, Irvine, CA 92697-2555 USA		Submitted (12-MAR-2001) Ecology and Evolutionary Biology, University of California, Irvine, 321 Steinhaus Hall, Irvine, CA 92697-2555 USA	
Location/Qualifiers		Location/Qualifiers	
1. 1290		1. 1290	
/organism="Trypanosoma cruzi"		/organism="Trypanosoma cruzi"	
/strain="MSC2"		/strain="MSC2"	
/db_xref="taxon:5693"		/db_xref="taxon:5693"	
<1. .>1290		<1. .>1290	
/gene="rR"		/gene="rR"	
<1. .>1290		<1. .>1290	
/gene="rR"		/gene="rR"	
/product="Trypanothione reductase"		/product="Trypanothione reductase"	
<1. .>1290		<1. .>1290	
/gene="rR"		/gene="rR"	
/codon_start=1		/codon_start=1	
/product="Trypanothione reductase"		/product="Trypanothione reductase"	
/protein_id="AAK1213.1"		/protein_id="AAK1213.1"	
/db_xref="GI:14600078"		/db_xref="GI:14600078"	
/translation="MLKRRVAVIDVNVHGPFSALGTCVNVGCVPKKLTATGACGSLKSRVAVNRKESADPASVAKERLEETENILASGSMPPHMPVPIETICISLEFLFGPEPRKRVLVGGFIEVERAGIENAKPKPDGVITLCYGDALIRFEDHTEELTSSNAFLVIRNGQVVDVDESRFNVSNIYVAGIGVNRVWLTVAINEAALVDVITFGTNPRTDHRVARSVAVSPIGTIGGLIEVASKREVAAYTSSFTVLMHNISGSKYKTFVAKITTNHSDGVTLGVHILGDAPRIITGTCICLAKNISDFY"		/translation="MLKRRVAVIDVNVHGPFSALGTCVNVGCVPKKLTATGACGSLKSRVAVNRKESADPASVAKERLEETENILASGSMPPHMPVPIETICISLEFLFGPEPRKRVLVGGFIEVERAGIENAKPKPDGVITLCYGDALIRFEDHTEELTSSNAFLVIRNGQVVDVDESRFNVSNIYVAGIGVNRVWLTVAINEAALVDVITFGTNPRTDHRVARSVAVSPIGTIGGLIEVASKREVAAYTSSFTVLMHNISGSKYKTFVAKITTNHSDGVTLGVHILGDAPRIITGTCICLAKNISDFY"	
309 a 318 c 353 g 310 e		309 a 318 c 353 g 310 e	
BASE COUNT		BASE COUNT	
ORIGIN		ORIGIN	
Query Match		Query Match	
Best Local Similarity 68.5%; Score 17.8; DB 3; Length 1290;		Best Local Similarity 68.5%; Score 17.8; DB 3; Length 1290;	
Matches 19; Conservativity 0; Mismatches 2; Indels 0; Gaps 0;		Matches 19; Conservativity 0; Mismatches 2; Indels 0; Gaps 0;	
1 CGCTCTTGCACATCGTACC 21		1 CGCTCTTGCACATCGTACC 21	
996 CGCTCTTGCACATCGTACC 976		996 CGCTCTTGCACATCGTACC 976	
RESULT 12		RESULT 12	
AF358997/c		AF358997/c	
LOCUS		LOCUS	
DEFINITION		DEFINITION	
AF358997		AF358997	
ACCESSION		ACCESSION	
VERSION		VERSION	
KEYWORDS		KEYWORDS	
SOURCE		SOURCE	
ORGANISM		ORGANISM	
Trypanosoma cruzi.		Trypanosoma cruzi.	
Trypanosoma cruzi		Trypanosoma cruzi	
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;		Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;	
Trypanosoma; Schizotrypanum.		Trypanosoma; Schizotrypanum.	
1 (bases 1 to 1290)		1 (bases 1 to 1290)	
Machado, C.A. and Ayala, F.J.		Machado, C.A. and Ayala, F.J.	
Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of <i>Trypanosoma cruzi</i>		Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of <i>Trypanosoma cruzi</i>	
Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001)		Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001)	
2 (bases 1 to 1290)		2 (bases 1 to 1290)	
Machado, C.A. and Ayala, F.J.		Machado, C.A. and Ayala, F.J.	
Direct Submission		Direct Submission	
Submitted (12-MAR-2001) Ecology and Evolutionary Biology, University of California, Irvine, 321 Steinhaus Hall, Irvine, CA 92697-2555 USA		Submitted (12-MAR-2001) Ecology and Evolutionary Biology, University of California, Irvine, 321 Steinhaus Hall, Irvine, CA 92697-2555 USA	
Location/Qualifiers		Location/Qualifiers	
1. 1290		1. 1290	
/organism="Trypanosoma cruzi"		/organism="Trypanosoma cruzi"	
/strain="MSC2"		/strain="MSC2"	
/db_xref="taxon:5693"		/db_xref="taxon:5693"	
<1. .>1290		<1. .>1290	
/gene="rR"		/gene="rR"	
<1. .>1290		<1. .>1290	
/gene="rR"		/gene="rR"	
/product="Trypanothione reductase"		/product="Trypanothione reductase"	
<1. .>1290		<1. .>1290	
/gene="rR"		/gene="rR"	
/codon_start=1		/codon_start=1	
/product="Trypanothione reductase"		/product="Trypanothione reductase"	
/protein_id="AAK1213.1"		/protein_id="AAK1213.1"	
/db_xref="GI:14600078"		/db_xref="GI:14600078"	

Machado C.A. and Ayala F.J.
Direct Submission
Submitted (12-MAR-2001) Ecology and Evolutionary Biology,
University of California, Irvine, 321 Steinhaus Hall, Irvine, CA
92697-2525, USA

FEATURES
SOURCE Location/Qualifiers
1. 1290
/organism="Trypanosoma cruzi"
/strain="SO3 c15"
/db_xref="taxon:5693"
/haplotype="2"
1. .>1290
/gene="rrr"
/gene="rrr"
1. .>1290
/product="trypanothione reductase"
1. .>1290
/gene="rrr"
/codon_start=1
/product="trypanothione reductase"
/protein_id="AAK71215.1"
/db_xref="GI:14600032"
/translation="TLKKRVAVIDQVHGPFFSALGTCVNVCVPKKILAMTYGAQ
GLSKIKVVNVRESDADRAVAKRLEETENILLASGSWPHMPPIPEICHSINAPPLI
PEPRRLVTGGGTIVSEFAGIRNAKPPDQGVTLCKRGMILRGFDITLLEETKL
TANGIOLITKENPAKVELNADSGSVTFESGKKMFEDLVMAAIGSPPTKLOLNQNG
VMINGGVQVDERTSRINSNIVAIQDVTRVNLTPVALINEAAIYDTVFGRPKTIQH
TVASAFVLPPIGTGGLIEVASRRVEVAVAYLSSTPILAHNTISGSKYTKFAKITT
NMSDYVLGHLLGDNAPEIIQVGICLKLNKISDFY"

CDS
mRNA

gene
CDS

Query Match
Best Local Similarity 68.5%; Score 17.8; DB 3; Length 1290;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 CGCTTTCCGACATCGGTACC 21
|||||
996 CGCTTTCCGACATCGGTACC 976

Db

RESULT 13
AF358999/c
LOCUS
DEFINITION Trypanosoma cruzi strain PSC-O haplotype 2 trypanothione reductase
ACCESSION AF358999
VERSION AF358999.1 GI:14600035
KEYWORDS
SOURCE
ORGANISM
TRYPANOSOMA CRUZI
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
1. (bases 1 to 1290)
Machado C.A. and Ayala F.J.
Nucleotide sequences provide evidence of genetic exchange among
distinctly related lineages of Trypanosoma cruzi
Proc. Natl. Acad. Sci. U.S.A. 98 (13), 7396-7401 (2001)
21310003
11416213
2 (bases 1 to 1290)
Machado C.A. and Ayala F.J.
Direct Submission
Submitted (12-MAR-2001) Ecology and Evolutionary Biology,
University of California, Irvine, 321 Steinhaus Hall, Irvine, CA
92697-2525, USA

FEATURES
SOURCE Location/Qualifiers
1. 1290
/organism="Trypanosoma cruzi"
/strain="PSC O"
/db_xref="taxon:5693"
/haplotype="2"

FEATURES	source	location/Qualifiers
gene	gene	1. >1290
mRNA	mRNA	1. >1290
CDS	CDS	1. >1290
ORIGIN	ORIGIN	1. >1290
BASE COUNT	BASE COUNT	309 a 318 c 353 g 310 t
ORIGIN	ORIGIN	1. >1290
BASE COUNT	BASE COUNT	309 a 318 c 353 g 310 t

Fri Jul 11 15:29:38 2003

us-10-087-631b-8.rge

Page 12

Query Match 68.5%; Score 17.8; DB 3; Length 1290;
Best Local Similarity 90.5%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGTCTTCGCGAGATCGGTACC 21
|||||
Db 996 CGTCTTCGCGAGATCGGTACC 976

Search completed: July 10, 2003, 19:20:30
Job time : 232.294 secs

Protophila melanog
Phycocephala abyss
CDNA of differentia
hOsf-2pl. Homo sa
DNA encoding Osf-2
TCI gene. Homo sa
Human cancer assoc
Human adenocarcino
Colon adenocarcino
hOsf-2os. Homo sa
DNA encoding Osf-2
Human Osf-2 (perio
Human cervical can
Human breast cancer
Mouse dectn-2 gene
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Aspergillus oryzae
Streptococcus pneu
Streptococcus pneu
Drosophila melanog
Streptococcus pneu
DNA encoding a hum
Human Immune/Naema
Human ovarian anti
Fusarium venenatu
Arabidopsis thalia
H. pylori cytoplast
Helicobacter pylori
Human secreted pro
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
cDNA sequence #139

PT Interactions

XX Claim 1; SEQ ID NO 41806; 21bp + Sequence listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB157737-AB172072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 768 BP; 173 A; 214 C; 218 G; 163 T; 0 other;

Query Match 67.7%; Score 17.6; DB 23; Length 768;
Best Local Similarity 83.3%; Pred. No. 49;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 2 GTCCTTCGACATCGTACTCA 25
600 GTCCTTCGCGGTCTACTCA 623

RESULT 2

AAC48208/c ID AAC48208 standard; DNA; 1215 BP.

AAC48208;

18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 56648.

XX Hybridization assay; genetic mapping; gene expression control;

XX Protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123348.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130891.

XX 28-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132485.

XX 07-MAY-1999; 99US-0132486.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134256.

XX 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134376.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 14-JUN-1999; 99US-0138847.

PR 16-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 17-JUN-1999; 99US-0139453.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139464.

PR 18-JUN-1999; 99US-0139465.

PR 18-JUN-1999; 99US-0139466.

PR 18-JUN-1999; 99US-0139467.

PR 18-JUN-1999; 99US-0139468.

PR 18-JUN-1999; 99US-0139469.

PR 18-JUN-1999; 99US-0139470.

PR 18-JUN-1999; 99US-0139471.

PR 18-JUN-1999; 99US-0140991.

PR 01-JUL-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 02-JUL-1999; 99US-0142154.

PR 06-JUL-1999; 99US-0142055.

PR 08-JUL-1999; 99US-0142390.

PR 09-JUL-1999; 99US-0142803.

PR 12-JUL-1999; 99US-0142920.

PR 13-JUL-1999; 99US-0142977.

PR 14-JUL-1999; 99US-0143542.

PR 15-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144634.

PR 21-JUL-1999; 99US-0144684.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 23-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.


```

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147040.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.

```

```

PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 67.7%; Score 17.6; DB 21; Length 1215;
 Best Local Similarity 83.3%; Pred. No. 52;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 3 TCTTCGACATCGCTACCTCAAT 26
DB 773 TCTTATCAGATGGAGACCTCAAT 750

```

RESULT 3
 ABL12431 standard; cDNA; 2112 BP.

XX ABL12431;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31775.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP. NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PSDB; ABB68328.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions

XX Claim 1; SEQ ID NO 31775; 21np + sequence listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 capable of detecting 1000 or more genes from Drosophila. The invention is
 useful in developmental biology and in elucidating cell signalling and
 cell-cell interactions in higher eukaryotes for the development of
 insecticides, therapeutics and pharmaceutical drugs. The invention
 discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 sequences (AB101840-AB16175) and the encoded proteins
 (AB157737-AB12072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2112 BP; 542 A; 555 C; 583 G; 432 T; 0 other;

Query Match 67.7%; Score 17.6; DB 23; Length 2112;
 Best Local Similarity 83.3%; Pred. No. 55;

ABL30110;
26-MAR-2002 (first entry)

AU9056291-A
18-JUN-1992

XX 05-JUN-1990; 90AU-0056291.
 PF Best Local Similarity 67.7%; Score 17.6; DB 13; Length 3522;
 XX 05-JUN-1990; 90AU-0056291.
 XX Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 PA (MYCO) MYCOGEN CORP.
 XX WPI: 1992-259364/32.
 DR P-PSDB: ABR25825.
 XX
 PT Controlling insect pests of lepidoptera family - includes
 PT contacting insects with suitable amt. of *Bacillus thuringiensis*
 PS Claim 13; Page 29; 49pp; English.
 XX
 CC The DNA sequence encodes a *Bacillus thuringiensis* endotoxin which
 CC may be used in an expression system to transform a variety of
 CC microbial hosts, e.g. *Pseudomonas*, *Azobacter*, *Erwinia*, *Serratia*,
 CC *Agrobacterium*, *Streptomyces* species, etc. Expression of the toxic
 CC gene results in the intracellular prodn. and maintenance of the
 CC peptide toxin. With suitable hosts, the microbes can be applied to
 CC the sites of lepidopteran insects, e.g. to the insects themselves,
 CC to the rhizosphere, phylloplane or to a body of water, where they
 CC will proliferate and be ingested by the insects. The result is the
 CC control of unwanted insects. Alternatively, the microbe hosting the
 CC toxin gene can be treated under conditions that prolong the
 CC activity of the toxin produced in the cell. The treated cell can
 CC be applied to the environment of target pests. The resulting prod.
 CC retains the toxicity of the B.t. toxin. See also AAQ26929-31.
 XX
 SQ Sequence 3522 BP; 1171 A; 579 C; 766 G; 1006 T; 0 other;
 OY
 Query Match 67.7%; Score 17.6; DB 13; Length 3522;
 Best Local Similarity 83.3%; Pred. No. 59;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 DB 3 TCTTTCGACATCGTACTCTCAAT 26
 252 TCTTTCGACATCGTACTCTCAAT 229
 RESULT 7
 ID ABL30114/c
 XX ABL30114 standard; DNA; 3704 BP.
 AC ABL30114;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE *Drosophila melanogaster* genomic polynucleotide SEQ ID NO 41815.
 XX
 KM *Drosophila*: developmental biology; cell signalling; insecticide;
 KM pharmaceutical; gene; ds.
 OS *Drosophila melanogaster*.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, LI PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell

PT interactions -
 XX
 PS Claim 1; SEQ ID NO 41815; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pt_sequences.
 XX
 SQ Sequence 3704 BP; 782 A; 963 C; 961 G; 998 T; 0 other;
 OY
 Query Match 67.7%; Score 17.6; DB 23; Length 3704;
 Best Local Similarity 83.3%; Pred. No. 59;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 DB 2 GTCCTTCGACATCGTACTCTCA 25
 3419 GTCCTTCGACATCGTACTCTCA 3396
 RESULT 8
 ID ABL12430/c
 XX ABL12430 standard; CDNA; 4233 BP.
 AC ABL12430;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE *Drosophila melanogaster* expressed polynucleotide SEQ ID NO 31772.
 XX
 KM *Drosophila*: developmental biology; cell signalling; insecticide;
 KM pharmaceutical; gene; ss.
 OS *Drosophila melanogaster*.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, LI PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR P-PSDB: ABB68327.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Claim 1; SEQ ID NO 31772; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
 CC sequences (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 4233 BP; 1150 A; 979 C; 934 G; 1170 T; 0 other;

Query Match
Best Local Similarity 67.7%; Score 17.6; DB 23; Length 4233;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

DB 2 GCTTTTCGAGATCGGTACTCTCAA 25
3009 GCTTTTCGAGATCGGTACTCTCAA 2986

RESULT 9
ABL13658/c
ID ABL13658 standard; cDNA; 8173 BP.

AC ABL13658;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 35456.
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN W0200171042-A2.

FD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

DR WPI; 2001-656860/75.
DR P-PSDB; ABB69555.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions

CC Claim 1; SEQ ID NO 35456; 21bp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 8173 BP; 2452 A; 1765 C; 1688 G; 2268 T; 0 other;

Query Match
Best Local Similarity 67.7%; Score 17.6; DB 23; Length 8173;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

DB 2 GCTTTTCGAGATCGGTACTCTCAA 25
204 GCTTTTCGAGATCGGTACTCTCAA 181

RESULT 10

ABL10262/c
ID ABL10262 standard; cDNA; 7407 BP.

AC ABL10262;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 25268.

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN W0200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

DR WPI; 2001-656860/75.
DR P-PSDB; ABB66159.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions

CC Claim 1; SEQ ID NO 25268; 21bp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 7407 BP; 1810 A; 1780 C; 1770 G; 2047 T; 0 other;

Query Match
Best Local Similarity 65.4%; Score 17; DB 23; Length 7407;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

DB 1 GCTTTTCGAGATCGGTACTCTCAA 25
313 GCTTTTCGAGATCGGTACTCTCAA 289

RESULT 11

AAH41225/c
ID AAH41225 standard; DNA; 349980 BP.

AC AAH41225;

DT 29-OCT-2001 (first entry)

DE Pyrococcus abyssi genomic fragment #4.

KW Hyperthermophilic archaeon; hyperthermophilic protein; ds.

Key	Location/Qualifiers
OS	Homo sapiens.
PH	Key
FT	CDS
XX	28..2367
XX	/*tag- a
XX	NO200035473-r2.
XX	22-JUN-2000.
XX	15-DEC-1999; 99MO-US29941.
XX	18-DEC-1998; 98US-0113008.
XX	(SCIO-) SCIOS INC.
XX	Stanton LM, White RT, Damm DL, Lewicki JA, Joly A, Schreiner GF;
XX	WPI, 2000-451904/39.
XX	P-PSDB; AAY93596.
XX	Preventing, diagnosing and treating cardiac, kidney and inflammatory
XX	disorders using cardiac genes, that are differentially expressed in
XX	disease states such as cardiac arrhythmia and arteriosclerosis -
XX	Disclosure; Fig 7C; 170pp; English.
XX	AAA46668-79 represent differentially expressed human genes, associated
XX	with disease states and disorders. The specification describes
XX	methods preventing, diagnosing and treating cardiac, kidney and
XX	inflammatory diseases associated with inappropriate expression of
XX	differentially expressed cardiac, kidney and inflammatory genes
XX	(e.g. AAA46668-79). These genes include I-8U, prostacyclin-stimulating
XX	factor, ISF-2, tissue specific mRNA, insulin-like growth factor
XX	binding protein 6, OSF-1, gas-1, YMP, ERG2, pre-B cell stimulating
XX	factor homologue (SDF1a), peripheral benzodiazepine receptor, and
XX	cellular ligand of annexin II (p11), respectively. These diseases
XX	include congenital heart failure, dilated congestive cardiomyopathy,
XX	hypertrophic cardiomyopathy, restrictive cardiomyopathy, mitral valve
XX	disease, aortic valve disease, tricuspid valve disease, angina pectoris,
XX	myocardial infarction, cardiac arrhythmia, pulmonary hypertension,
XX	arterial hypertension, renovascular hypertension, arteriosclerosis,
XX	atherosclerosis and/or cardiac tumours.
XX	Sequence 3077 BP; 1032 A; 571 C; 632 G; 842 T; 0 other;
XX	Query Match 63.8%; Score 16.6; DB 21; Length 3077;
XX	Best Local Similarity 82.6%; Pred. No. 1.8e+02;
XX	Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX	4 CTTTCGCAGATCGGTACTCTCAAT 26
XX	
XX	715 CTTACACAAATGCTACTCTCAAT 737
XX	RESULT 13
XX	AA048934
XX	AA048934 standard; cDNA to mRNA; 3092 BP.
XX	AA048934;
XX	15-APR-1994 (first entry)
XX	hOSF-2pL.
XX	Bone-related protein; bone; diagnosis; disease; growth factor;
XX	cell adhesion; guiding; induction; metabolic bone disease; ss.

```

XX OS Homo sapiens (placenta).
XX XX
XX Key Location/Qualifiers
XX FH CDS 38..2377
XX FT sig_peptide /tag= a
XX FT sig_peptide /product= OSF-2
XX FT mat_peptide /tag= b
XX FT mat_peptide 102..2374
XX FT /tag= c
XX PN EP62508-A.
XX XX
XX PD 29-SEP-1993.
XX PF 22-MAR-1993; 93EP-0104650.
XX PR 27-MAR-1992; 92JP-0071501.
XX PA (FARH ) HOECHST JAPAN LTD.
XX PI Anann E, Kikuno R, Otawara-Hamamoto Y, Takeshita S;
XX PI Tezuka K;
XX DR WPI; 1993-304910/39.
XX DR P-PSDB; AAR41868.
XX XX
XX PT Mammalian OSF-2 protein - obtd. from bone tissue, useful for
XX XX diagnosis and treatment of metabolic bone disease
XX PS Claim 6; Page 14-18; 58pp; English.
XX CC OSF-2 plays an important role in the formation of bone, by acting as
XX CC a growth factor or adhesion or "guiding" protein to attract cells
XX CC to the site of bone induction. In a pharmaceutical compn. it can
XX CC be applied in metabolic bone diseases.
XX SQ Sequence 3092 BP; 1035 A; 573 C; 638 G; 846 T; 0 other;

Query Match 63.8%; Score 16.6; DB 14; Length 3092;
Best Local Similarity 82.6%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 CTTTCGAGATCGTACTCAAT 26
DB 725 CTTACACAAATTGTACTCAAT 747

RESULT 14
ABK49493 standard; DNA; 3092 BP.
XX AC ABK49493;
XX XX
XX DT 15-JUL-2002 (first entry)
XX XX
XX DE DNA encoding OSF-2 protein associated protein #3.
XX XX
XX KW Heart failure; OSF-2 protein; gene therapy; transgenic animal;
XX KW human; gene; ds.
XX OS Homo sapiens.
XX XX
XX FT Key Location/Qualifiers
XX FT CDS 38..2377
XX FT /tag= a
XX FT /product= "OSF-2 associated protein #3"
XX PN W0200220055-A1.
XX PD 14-MAR-2002.
XX XX

```

```

PF 07-SEP-2001; 2001WO-JP07787.
XX XX
XX PR 08-SEP-2000; 2000JP-0273644.
XX XX
XX PA (SUNR ) SUNTORI LTD.
XX PA (SUNT-) SUNTORI BIOMEDICAL RES LTD.
XX XX
XX PI Kawashima K, Katsuragi N, Sugimura K, Furuya M, Morishita R;
XX DR WPI; 2002-371858/40.
XX DR P-PSDB; AAT79825.
XX XX
XX PT Remedies for heart failure obtained by screening drugs capable of
XX PT inhibiting expression of OSF-2 gene or production or function of the
XX PT protein -
XX PS Claim 5; Page 61-66; 80pp; Japanese.
XX XX
XX CC The invention describes preventives or remedies for heart failure
XX CC comprising a substance inhibiting the expression of an OSF (undefined)-2
XX CC protein or function of the protein or function of a target molecule of
XX CC the OSF-2 protein. The remedies are for heart failure, and are also
XX CC applicable for gene therapy. Diagnosis can also be achieved by monitoring
XX CC gene expression and protein production. Transgenic animals are useful for
XX CC studying the degree of disease deterioration and onset risk forecast to
XX CC provide treatment and help adjust lifestyle. This sequence encodes an
XX CC OSF-2 associated protein.
XX SQ Sequence 3092 BP; 1035 A; 573 C; 638 G; 846 T; 0 other;

Query Match 63.8%; Score 16.6; DB 24; Length 3092;
Best Local Similarity 82.6%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 CTTTCGAGATCGTACTCAAT 26
DB 725 CTTACACAAATTGTACTCAAT 747

RESULT 15
AAQ90112 standard; CDNA; 3126 BP.
XX ID AAQ90112;
XX AC AAQ90112;
XX XX
XX DT 05-NOV-1995 (first entry)
XX XX
XX DE TCI gene.
XX XX
XX KW Tumour marker; invasive; metastatic; cancer; ss.
XX KW Homo sapiens.
XX OS
XX FH Key Location/Qualifiers
XX FH CDS 43..2376
XX FT /tag= a
XX FT
XX PN W09511923-A.
XX PN
XX PD 04-MAY-1995.
XX XX
XX PF 31-OCT-1994; 94WO-US12502.
XX PR 29-OCT-1993; 93US-0146488.
XX PA (DAND ) DANA FARMER CANCER INST INC.
XX PA Bao S, Chen LB, Liu Y;
XX PI
XX DR WPI; 1995-176826/23.
XX DR P-PSDB; AAR74302.
XX XX
XX PT New tumour marker TCI, corresp. DNA and monoclonal antibody - for

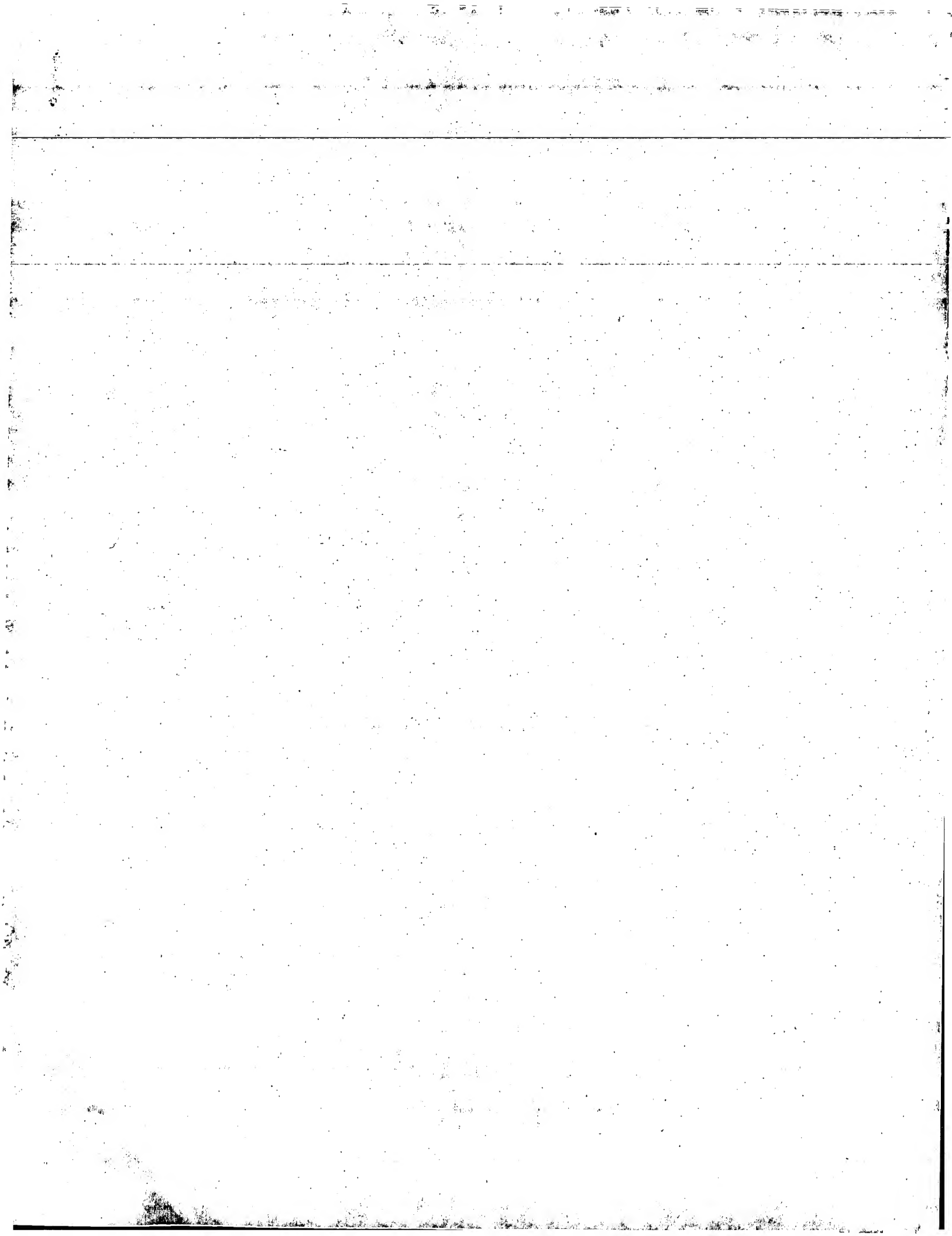
```

XX Disclosure; Flg 4; 84pp; English.

SQ Sequence 3126 BP; 1035 A; 611 C; 656 G; 824 T; 0 other;

4 CTTTCGAGATCGGTACTCAAT 26

Search completed: July 10, 2003, 19:52:45
Job time : 57.8629 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 ; Search time 11.627 Seconds

(Without alignments)
685 / 785 Million cell updates/sec

Title: US-10-087-631b-8

Perfect score: 26

Sequence: 1 cgccttcgcagcgcgtactcaat 26

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:**

2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:**

3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:**

4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:**

5: /cgn2_6/ptodata/1/ina/PCUTS.COMB.seq:**

6: /cgn2_6/ptodata/1/ina/Backfillseq:**

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	65.4	495	4	US-09-221-017B-1098
2	16.6	63.8	636	2	US-08-477-396A-1
3	16.6	63.8	3092	1	US-08-426-627-3
4	16.6	63.8	3126	2	US-08-477-396A-3
5	16.6	63.8	3253	1	US-08-426-627-5
6	16.6	63.8	10409	3	US-08-772-440-33
7	16.4	63.1	2502	2	US-08-844-086-1
8	16.4	63.1	2502	2	US-09-018-211-1
9	16.4	63.1	11770	4	US-08-961-527-172
10	16	61.5	2373	4	US-08-975-762-45
11	16	61.5	2373	4	US-09-295-028-45
12	16	61.5	2373	4	US-09-106-582-45
13	16	61.5	2574	4	US-09-253-829-28
14	16	61.5	3522	1	US-08-040-751-4
15	16	61.5	3522	1	US-08-291-368-1
16	16	61.5	3522	2	US-08-962-190-1
17	16	61.5	3522	5	PCT-US95-10310-1
18	16	61.5	3522	6	5164180-3
19	16	61.5	4003	1	US-08-257-999-1
20	16	61.5	7195	4	US-08-478-507-6
21	16	61.5	7195	4	US-09-128-275A-6
22	16	61.5	7195	4	US-09-553-427-6
23	15.6	60.0	4635	4	US-09-221-017B-187
24	15.4	59.2	1200	4	US-09-502-653-7
25	15.4	59.2	1875	2	US-08-683-743-3
26	15.4	59.2	99500	4	US-09-798-096-10
27	15.2	58.5	44	1	US-08-344-695-5

28	15.2	58.5	580	4	US-09-328-111-465	Sequence 465, App
29	15.2	58.5	1224	1	US-07-718-274A-5	Sequence 5, Appl
30	15.2	58.5	1224	1	US-08-149-106-5	Sequence 5, Appl
31	15.2	58.5	1224	1	US-08-298-021-5	Sequence 5, Appl
32	15.2	58.5	1586	1	US-07-841-646-18	Sequence 18, Appl
33	15.2	58.5	1586	1	US-08-147-023-18	Sequence 18, Appl
34	15.2	58.5	1586	1	US-08-447-570-18	Sequence 18, Appl
35	15.2	58.5	1586	2	US-08-449-700-18	Sequence 18, Appl
36	15.2	58.5	1586	2	US-08-449-699A-18	Sequence 18, Appl
37	15.2	58.5	1751	1	US-08-377-292-4	Sequence 4, Appl
38	15.2	58.5	1788	1	US-07-841-646-6	Sequence 6, Appl
39	15.2	58.5	1788	1	US-07-901-703-18	Sequence 18, Appl
40	15.2	58.5	1788	1	US-08-147-023-6	Sequence 6, Appl
41	15.2	58.5	1788	1	US-08-447-570-6	Sequence 6, Appl
42	15.2	58.5	1788	2	US-08-449-700-6	Sequence 6, Appl
43	15.2	58.5	1788	2	US-08-449-699A-6	Sequence 6, Appl
44	15.2	58.5	1788	5	PCT-US93-05446-18	Sequence 18, Appl
45	15.2	58.5	1954	1	US-08-050-132A-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-221-017B-1098
Sequence 1098, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: ROSS, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq. for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 1098:
SEQUENCE CHARACTERISTICS:
LENGTH: 495 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE: PORYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc-feature
LOCATION: 1...495
US-09-221-017b-1098

Query Match 65.4%; Score 17; DB 4; Length 495;
Best Local Similarity 80.0%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1 CGCTCTTCGACATCGTACTCTCA 25
228 CGACTGTACGATCGTACTCTCA 252

RESULT 2

US-08-477-396A-1
Sequence 1, Application US/08477396A
Patent No. 5872235

GENERAL INFORMATION:

APPLICANT: Chen, Ian Bo
APPLICANT: Bao, Shideng
APPLICANT: Liu, Yuan
TITLE OF INVENTION: A NOVEL TUMOR MARKER AND NOVEL METHOD OF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weingarten, Schungtin, Gagnebin & Hayes
STREET: Ten Post Office Square
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,396A
FILING DATE:

CLASSIFICATION:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/146,488

FILING DATE: 29-OCT-1993

APPLICATION NUMBER: US 08/448,388

FILING DATE: 28-MAY-1996

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/12502

FILING DATE: 31-OCT-1994

ATTORNEY/AGENT INFORMATION:
NAME: Helene, Holliday C.

REGISTRATION NUMBER: 34,346

REFERENCE/DOCKET NUMBER: DFCI-333BX

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-2290

TELEFAX: (617) 451-0313

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 636 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

FEATURE:
NAME/KEY: CDS
LOCATION: 1..636

US-08-477-396A-1

Query Match 63.8%; Score 16.6; DB 2; Length 636;
Best Local Similarity 82.6%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 4 CTTTCGACATCGTACTCTCAAT 26
58 CTTACACAAATTTGTACTCTCAAT 80

RESULT 3

US-08-426-627-3
Sequence 3, Application US/08426627
Patent No. 575664

GENERAL INFORMATION:

APPLICANT: Amann, Egon
APPLICANT: Otawara-Hamamoto, Yoko
APPLICANT: Kikuno, Reiko
APPLICANT: Takeshita, Sunao
APPLICANT: Tezuka, Kenichi
TITLE OF INVENTION: NO. 575664e1 Protein with Bone Formation
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/426,627
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/036,841

FILING DATE: 25-MAR-1993

APPLICATION NUMBER: JP 4-71501

FILING DATE: 27-MAR-1992

ATTORNEY/AGENT INFORMATION:

NAME: Hammond, Alan W.

REGISTRATION NUMBER: 35,178

REFERENCE/DOCKET NUMBER: 02481-1285-00000

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4400

TELEFAX: 202-408-4400

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 3092 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

MOLECULE TYPE: cDNA to mRNA

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

TISSUE TYPE: Placenta

FEATURE:
NAME/KEY: CDS
LOCATION: join(38..2375)

FEATURE:
NAME/KEY: mat-peptide
LOCATION: join(101..2375)

US-08-426-627-3

Query Match 63.8%; Score 16.6; DB 1; Length 3092;
Best Local Similarity 82.6%; Pred. No. 29;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CTTTGGCAGATGGTACTCAAT 26

DB 725 CTTACACAATTTGTACTCAAT 747

RESULT 4

US-08-477-396A-3

; Sequence 3, Application US/08477396A

; Patent No. 5872235

; GENERAL INFORMATION:

; APPLICANT: Chen, Ian Bo

; APPLICANT: Bao, Shideng

; APPLICANT: Liu, Yuan

; TITLE OF INVENTION: A NOVEL TUMOR MARKER AND NOVEL METHOD OF

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Weingarten, Schurgin, Gagnabin & Hayes

; STREET: Ten Post Office Square

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/477,396A

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/146,488

; FILING DATE: 29-OCT-1993

; APPLICATION NUMBER: US 08/448,388

; FILING DATE: 28-MAY-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/12502

; FILING DATE: 31-OCT-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Heine, Holliday C.

; REGISTRATION NUMBER: 34,346

; REFERENCE/DOCKET NUMBER: DFCI-333BX

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-2290

; TELEFAX: (617) 451-0313

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3126 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 43..2376

; US-08-477-396A-3

Query Match 63.8%; Score 16.6; DB 2; Length 3126;

Best Local Similarity 82.6%; Pred. No. 29;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CTTTGGCAGATGGTACTCAAT 26

DB 748 CTTACACAATTTGTACTCAAT 770

RESULT 5

US-08-426-627-5

; Sequence 5, Application US/08426627

; Patent No. 5756664

; GENERAL INFORMATION:

; APPLICANT: Amano, Egon

; APPLICANT: Otawara-Hamamoto, Yoko

; APPLICANT: Kikuno, Reiko

; APPLICANT: Takeshita, Sunao

; APPLICANT: Tezuka, Kenichi

; TITLE OF INVENTION: No. 5756664e1 Protein with Bone Formation

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finegan, Henderson, Farabow, Garrett &

; ADDRESS: Dunner

; STREET: 1300 I Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/426,627

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/036,841

; FILING DATE: 25-MAR-1993

; APPLICATION NUMBER: JP 4-71501

; FILING DATE: 27-MAR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Hammond, Alan W.

; REGISTRATION NUMBER: 35,178

; REFERENCE/DOCKET NUMBER: 02481-1285-00000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-408-4000

; TELEFAX: 202-408-4400

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3253 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; CELL TYPE: osteosarcoma

; FEATURE:

; NAME/KEY: CDS

; LOCATION: join(32..2540)

; FEATURE:

; NAME/KEY: mat.peptide

; LOCATION: join(97..2540)

; US-08-426-627-5

Query Match 63.8%; Score 16.6; DB 1; Length 3253;

Best Local Similarity 82.6%; Pred. No. 29;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CTTTGGCAGATGGTACTCAAT 26

DB 719 CTTACACAATTTGTACTCAAT 741

RESULT 6

US-08-772-440-33

; Sequence 33, Application US/08772440

; Patent No. 6046158

; GENERAL INFORMATION:

APPLICANT: Arizumi, Kiyoshi.
APPLICANT: Takashima, Akira
TITLE OF INVENTION: UNIQUE DENDRITIC CELL-ASSOCIATED C-TYPE
TITLE OF INVENTION: LECTINS, DECTIN-1 AND DECTIN-2; COMPOSITIONS AND USES
TITLE OF INVENTION: THEROP
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESS: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/772,440
FILING DATE: CONCURRENTLY HEREWITH
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTXD:493
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 10409 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: modified_base
LOCATION: 6510
OTHER INFORMATION: /mod_base= OTHER
OTHER INFORMATION: /note= "D = A or G or T"
FEATURE:
NAME/KEY: modified_base
LOCATION: 3406..6470
OTHER INFORMATION: /mod_base= OTHER
OTHER INFORMATION: /note= "K = G or T"
FEATURE:
NAME/KEY: modified_base
LOCATION: 3564..7896
OTHER INFORMATION: /mod_base= OTHER
OTHER INFORMATION: /note= "M = A or C"
FEATURE:
NAME/KEY: modified_base
LOCATION: 3479..6422
OTHER INFORMATION: /mod_base= OTHER
OTHER INFORMATION: /note= "R = A or G"
FEATURE:
NAME/KEY: modified_base
LOCATION: 3405..6871
OTHER INFORMATION: /mod_base= OTHER
OTHER INFORMATION: /note= "S = C or G"
FEATURE:
NAME/KEY: modified_base
LOCATION: 3457..9958
OTHER INFORMATION: /mod_base= OTHER
OTHER INFORMATION: /note= "W = A or T"
FEATURE:
NAME/KEY: modified_base
LOCATION: 3595..9959

OTHER INFORMATION: /mod_base= OTHER
OTHER INFORMATION: /note= "Y = C or T"
US-08-772-440-33
Query Match
Best Local Similarity 79.2%; Score 16.6; DB 3; Length 10409;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 3 TCTTGGCAGATCGCTACTCAAT 26
DB 2852 TCTTGGCAGATCGCTACTCAAT 2875
RESULT 7
US-08-844-086-1
Sequence 1, Application US/08844086
Patent No. 5866390
GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 5866390e1 Compounds
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Smithline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fastseq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,086
FILING DATE: 18-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimml, Edward R.
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2502 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-844-086-1
Query Match
Best Local Similarity 76.9%; Score 16.4; DB 2; Length 2502;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CGCTTTGGCAGATCGCTACTCAAT 26
DB 590 CGCTTTGGCAGATCGCTACTCAAT 615
RESULT 8
US-09-018-211-1
Sequence 1, Application US/09018211
Patent No. 6048716
GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 6048716e1 Compounds

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,211
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/844,086
FILING DATE: 18-APR-1997
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2502 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-018-211-1

Query Match 63.1%; Score 16.4; DB 3; Length 2502;
Best Local Similarity 76.9%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CGCTTTGCGAGATCGTACTCAAT 26
DB 590 CGCCTTACGACGACGCTTCTCAAT 615

RESULT 9
US-08-961-527-172
Sequence 172; Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 172:
SEQUENCE CHARACTERISTICS:
LENGTH: 11770 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-172

Query Match 63.1%; Score 16.4; DB 4; Length 11770;
Best Local Similarity 76.9%; Pred. No. 44;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CGCTTTGCGAGATCGTACTCAAT 26
DB 7818 CGCCTTACGACGACGCTTCTCAAT 7843

RESULT 10
US-08-975-762-45
Sequence 45; Application US/08975762
Patent No. 6207169
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,762
FILING DATE: 21-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 2373 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-975-762-45

Query Match 61.5%; Score 16; DB 4; Length 2373;
Best Local Similarity 79.2%; Pred. No. 55;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 CGCTTTGCGAGATCGTACTCA 24

Db 910 CATCTTTCGCGATAGTACTCA 933

RESULT 11

US-09-295-028-45
Sequence 45, Application US/09295028
Patent No. 6277381

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
FILE REFERENCE: 210121.439C4

CURRENT APPLICATION NUMBER: US/09/295,028

NUMBER OF SEQ ID NOS: 85

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO: 45

LENGTH: 2373

TYPE: DNA

ORGANISM: Ehrlichia sp.

US-09-295-028-45

Query Match
Best Local Similarity 79.2%; Score 16; DB 4; Length 2373;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

RESULT 12

US-09-106-582-45
Sequence 45, Application US/09106582
Patent No. 6306402

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,582
FILING DATE: 29-JUN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ. ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 2373 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-106-582-45

Query Match
Best Local Similarity 79.2%; Score 16; DB 4; Length 2373;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 910 CATCTTTCGCGATAGTACTCA 933

RESULT 13

US-09-255-829-28/c
Sequence 28, Application US/09255829
Patent No. 6461617

GENERAL INFORMATION:

APPLICANT: Shone, Clifford Charles
APPLICANT: Quinn, Conrad Padraig
APPLICANT: Foster, Keith Alan
TITLE OF INVENTION: Recombinant Toxin Fragments
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/255,829
FILING DATE: 23-FEB-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB97/02273
FILING DATE: 22-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/782,893
FILING DATE: 27-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: ESMOND, ROBERT W.
REGISTRATION NUMBER: 32,893
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2540
TELEFAX: 202-371-2600
INFORMATION FOR SEQ. ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 2574 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2574
US-09-255-829-28

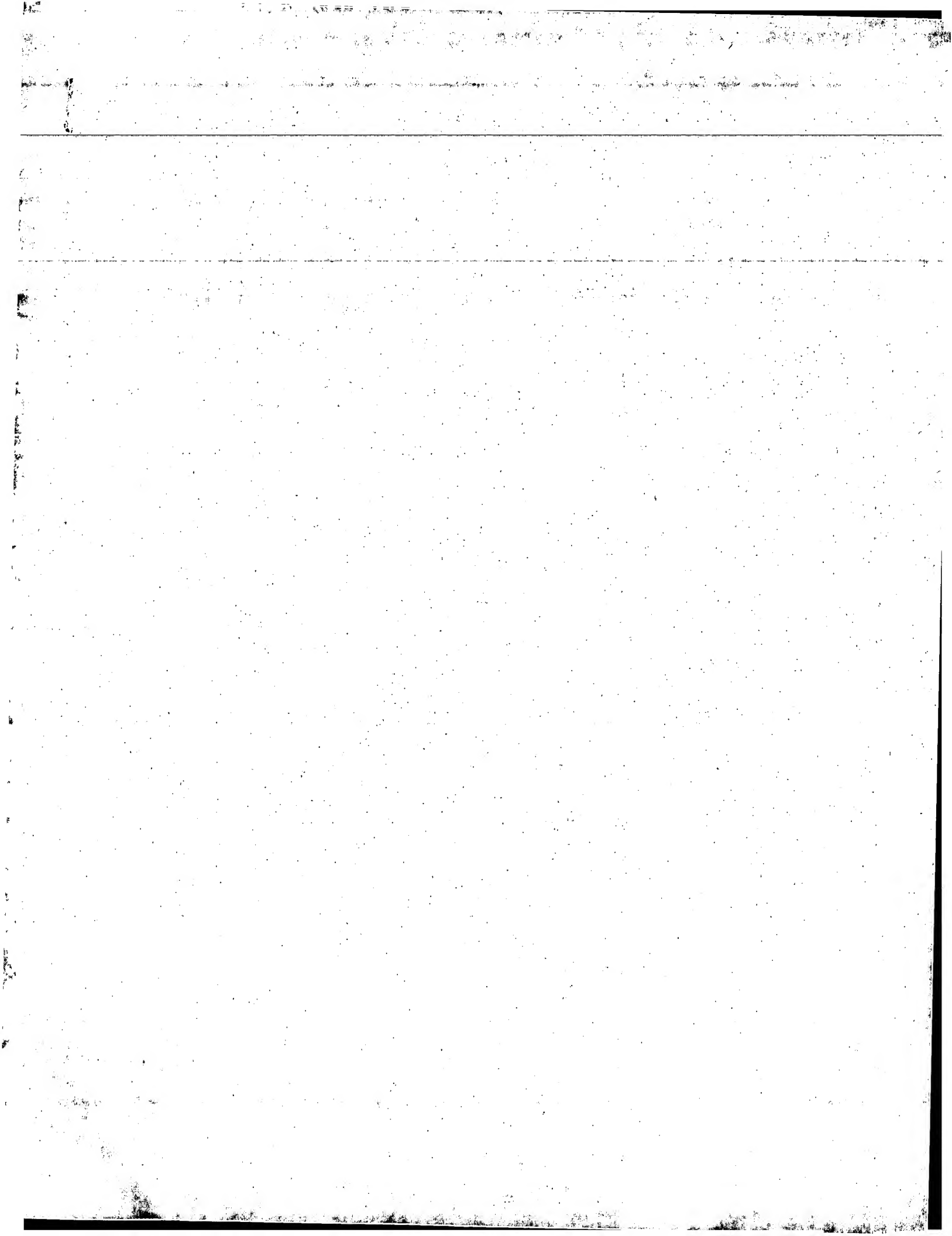
Query Match
Best Local Similarity 79.2%; Score 16; DB 4; Length 2574;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1570 GTTTTGTAGACCGGAGCTCAA 1547

RESULT 14

US-08-040-751-4/c
Sequence 4, Application US/08040751
Patent No. 5407825
GENERAL INFORMATION:
APPLICANT: Payne, Jewel M
APPLICANT: Sick, August J
TITLE OF INVENTION: No. 5407825el Bacillus thuringiensis isolates
TITLE OF INVENTION: active against Lepidopteran Pests and Genes Encoding No. 54078
TITLE OF INVENTION: Lepidopteran-active Toxins
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: DAVID R. SALIWANCHIK
STREET: 2421 N.W. 41st STREET, SUITE A-1
CITY: GAINESVILLE
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/040.751
FILING DATE: 19930329
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SALIWANCHIK, DAVID R.
REFERENCE/DOCKET NUMBER: MA39.C1.D3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3522 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
STRAIN: aizawai
INDIVIDUAL ISOLATE: PS81A2
IMMEDIATE SOURCE:
LIBRARY: Lambdagem - 11 (tm) Library of August Sick
CLONE: 81A2
US-08-040-751-4
Query Match 61.5%; Score 16; DB 1; Length 3522;
Best Local Similarity 79.2%; Pred. No. 58;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 3 TCTTTCGAGATCGGTACTCAAT 26
||||| 11 1111 11 1111111
DB 252 TCTTGGCGATCAATAGCTCAAT 229
RESULT 15
US-08-291-368-1/c
Sequence 1, Application US/08291368
Patent No. 5686069
GENERAL INFORMATION:
APPLICANT: Payne, Jewel M.
APPLICANT: Sick, August J.
TITLE OF INVENTION: No. 5686069el Bacillus thuringiensis Isolates
TITLE OF INVENTION: Active Against Lepidopteran Pests
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SalIWanchik & SalIWanchik
STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/291.368
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/597,607
FILING DATE: 15-OCT-90
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: SalIWanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA50.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904)375-8100
TELEFAX: (904)372-5800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3522 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
STRAIN: aizawai
INDIVIDUAL ISOLATE: PS81A2
IMMEDIATE SOURCE:
LIBRARY: Lambdagem - 11 (tm) Library of August Sick
CLONE: 81A2
US-08-291-368-1
Query Match 61.5%; Score 16; DB 1; Length 3522;
Best Local Similarity 79.2%; Pred. No. 58;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 3 TCTTTCGAGATCGGTACTCAAT 26
||||| 11 1111 11 1111111
DB 252 TCTTGGCGATCAATAGCTCAAT 229
Search completed: July 10, 2003, 20:27:41
Job time : 13.627 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: July 10, 2003, 22:53:16 ; Search time 71.164 Seconds
(without alignments)
576.891 Million cell updates/sec

Title: US-10-087-631B-8
Perfect score: 26
Sequence: 1 cgtcttcgcagatcgctacccaat 26

Scoring table: IDENTITY NUC
Gapop 10.0 ; Gapext 1.0

Searched: 1105431 seqs, 789497651 residues
Total number of hits satisfying chosen parameters: 2210662

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/PCCTUS_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	26	US-10-087-631B-8	Sequence 8, Appl1
2	24.4	99.8	241	US-10-087-631B-12	Sequence 12, Appl1
3	24.4	99.8	943	US-10-087-631B-11	Sequence 11, Appl1
4	17	65.4	25	US-10-088-263B-53770	Sequence 53770, A
5	16.6	63.8	507	US-09-796-692-4795	Sequence 4795, Ap
6	16.6	63.8	507	US-10-040-862-4795	Sequence 4795, Ap
7	16.6	63.8	2190	US-10-217-371-5	Sequence 13, Appl1
8	16.6	63.8	2253	US-10-217-371-13	Sequence 13, Appl1
9	16.6	63.8	2253	US-10-217-371-3	Sequence 9, Appl1
10	16.6	63.8	2274	US-10-217-371-9	Sequence 11, Appl1
11	16.6	63.8	2313	US-10-217-371-11	Sequence 11, Appl1
12	16.6	63.8	2337	US-10-217-371-7	Sequence 1, Appl1
13	16.6	63.8	2508	US-10-217-371-1	Sequence 45, Appl1
14	16.6	63.8	2937	US-10-176-847-45	Sequence 179, App
15	16.6	63.8	3077	US-10-171-311-119	Sequence 177, App
16	16.6	63.8	3202	US-09-925-301-171	Sequence 341, App
17	16.6	63.8	3213	US-10-171-311-177	Sequence 9861, Ap
18	16.6	63.8	3213	US-10-177-293-341	
19	16.6	63.8	3681	US-10-198-846-9861	

20	16.6	63.8	397658	10	US-09-813-320-3	Sequence 3, Appl1
21	16.6	63.8	9025608	9	US-10-156-761-1	Sequence 1, Appl1
22	16.4	63.1	617	10	US-09-770-149-820	Sequence 820, App
23	16.4	63.1	1794	10	US-09-815-242-9128	Sequence 9128, Ap
24	16.4	63.1	5289	9	US-10-128-714-179	Sequence 179, App
25	16.4	63.1	5289	9	US-10-128-714-5179	Sequence 5179, Ap
26	16.6	61.5	25	9	US-10-215-112-6258	Sequence 6258, Ap
27	16.6	61.5	168	10	US-09-878-574-13775	Sequence 13775, A
28	16.6	61.5	490	9	US-09-918-995-33372	Sequence 33372, A
29	16.6	61.5	614	10	US-09-770-149-838	Sequence 838, App
30	16.6	61.5	705	10	US-09-815-242-7385	Sequence 7385, Ap
31	16.6	61.5	945	9	US-10-084-546-5	Sequence 5, Appl1
32	16.6	61.5	1410	9	US-09-938-842A-818	Sequence 818, App
33	16.6	61.5	1794	9	US-09-822-846-139	Sequence 139, App
34	16.6	61.5	2184	10	US-09-801-368-295	Sequence 295, App
35	16.6	61.5	2373	10	US-09-159-469-45	Sequence 45, Appl1
36	16.6	61.5	2373	10	US-09-798-042-45	Sequence 45, Appl1
37	16.6	61.5	3120	9	US-10-128-714-261	Sequence 261, App
38	16.6	61.5	3120	9	US-10-128-714-5261	Sequence 5261, App
39	16.6	61.5	3969	9	US-10-177-293-386	Sequence 386, App
40	16.6	61.5	4235	9	US-10-198-846-10905	Sequence 10905, A
41	16.6	61.5	4416	9	US-10-128-714-395	Sequence 395, App
42	16.6	61.5	4417	9	US-10-128-714-5395	Sequence 5395, App
43	16.6	61.5	7195	9	US-09-851-410-6	Sequence 6, Appl1
44	15.8	60.8	135	10	US-09-878-574-7595	Sequence 7595, App
45	15.8	60.8	333	10	US-09-294-093B-124	Sequence 124, App

ALIGNMENTS

RESULT 1
US-10-087-631B-8
Sequence 8, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
TITLE OF INVENTION: CONTROL
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087, 631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 8
LENGTH: 26
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: 5780pc primer parallel
US-10-087-631B-8

Query Match 100.0%; Score 26; DB 9; Length 26;
Best Local Similarity 100.0%; Pred.No. 0.0014;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCTCTTCGAGATCGGCTACCTCAAT 26
DB 1 CGCTCTTCGAGATCGGCTACCTCAAT 26
|||||

RESULT 2
US-10-087-631B-12
Sequence 12, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
TITLE OF INVENTION: CONTROL
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087, 631B
CURRENT FILING DATE: 2002-03-01

NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 12
LENGTH: 241
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Amplicon derived from OS (pc)
US-10-087-631B-12

Query Match
Best Local Similarity 93.8%; Score 24.4; DB 9; Length 241;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db
1 CGCTCTTCGACGATCGTACTCCTCAAT 26
1 CGCTCTTCGACGATCGTACTCCTCAAT 26

RESULT 3.
US-10-087-631B-11
Sequence 11, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 11
LENGTH: 943
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: OS (pc) HCV being parallel-
US-10-087-631B-11

Query Match
Best Local Similarity 93.8%; Score 24.4; DB 9; Length 943;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db
1 CGCTCTTCGACGATCGTACTCCTCAAT 26
57 CGCTCTTCGACGATCGTACTCCTCAAT 82

RESULT 4
US-10-098-263B-53770/c
Sequence 53770, Application US/10098263B
Publication No. US20030104410A1
GENERAL INFORMATION:
APPLICANT: Mittman, Michael
TITLE OF INVENTION: Human Microarray
FILE REFERENCE: 3118.1
CURRENT APPLICATION NUMBER: US/10/098,263B
CURRENT FILING DATE: 2003-01-08
PRIOR APPLICATION NUMBER: 60/276,759
NUMBER OF SEQ ID NOS: 131066
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
SEQ ID NO 53770
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapien
US-10-098-263B-53770

Query Match
Best Local Similarity 65.4%; Score 17; DB 9; Length 25;
Matches 80.0%; Pred. No. 46; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db
2 GCTCTTCGACGATCGTACTCCTCAAT 26
25 GTACTTCGCGGATGAGTACTCCTCAAT 1

RESULT 5
US-09-796-692-4795
Sequence 4795, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
FILE REFERENCE: 2077.001200
CURRENT APPLICATION NUMBER: US/09/796,692
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: 60/223,378
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 9597
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4795
LENGTH: 507
TYPE: DNA
ORGANISM: Homo sapiens
US-09-796-692-4795

Query Match
Best Local Similarity 63.8%; Score 16.6; DB 9; Length 507;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db
4 CTTTCGACGATCGTACTCCTCAAT 26
438 CTTACACAAATGTGACTCCTCAAT 460

RESULT 6
US-10-040-862-4795
Sequence 4795, Application US/10040862
Publication No. US20030078396A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
APPLICANT: Retter, Marc
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther
TITLE OF INVENTION: Hematological Malignancies

```
FILE REFERENCE: 014058-013520US
CURRENT APPLICATION NUMBER: US/10/040,862
CURRENT FILING DATE: 2001-11-06
PRIOR APPLICATION NUMBER: US 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: US 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: US 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: US 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: US 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: US 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: US 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: US 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: US 60/223,378
PRIOR FILING DATE: 2000-08-07
PRIOR APPLICATION NUMBER: US 09/796,692
PRIOR FILING DATE: 2001-03-01
NUMBER OF SEQ ID NOS: 10467
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4795
LENGTH: 507
TYPE: DNA
ORGANISM: Homo sapiens
US-10-040-862-4795

Query Match          63.8%; Score 16.6; DB 9; Length 507;
Best Local Similarity 82.6%; Pred. No. 94;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      4 CTTGCAGATCGTACTCCTCAAT 26
         ||| ||| ||| ||| ||| ||| |||
DB      438 CTTACACAATTGTACTCCTCAAT 460

RESULT 7
US-10-217-371-5
; Sequence 5, Application US/10217371
; Publication No. US20030073137A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Ian Bo
; APPLICANT: Dai, Meiru
; APPLICANT: Sasaki, Hidefumi
; APPLICANT: Auciailt, Daniel
; TITLE OF INVENTION: PERIOSTIN-BASED DIAGNOSTIC ASSAYS
; FILE REFERENCE: 00530-099001
; CURRENT APPLICATION NUMBER: US/10/217,371
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US 60/312,123
; PRIOR FILING DATE: 2001-08-13
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 2190
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2190)
US-10-217-371-5

Query Match          63.8%; Score 16.6; DB 9; Length 2190;
```

```
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      4 CTTGCAGATCGTACTCCTCAAT 26
         ||| ||| ||| ||| ||| ||| |||
DB      625 CTTACACAATTGTACTCCTCAAT 647

RESULT 8
US-10-217-371-13
; Sequence 13, Application US/10217371
; Publication No. US20030073137A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Ian Bo
; APPLICANT: Dai, Meiru
; APPLICANT: Sasaki, Hidefumi
; APPLICANT: Auciailt, Daniel
; TITLE OF INVENTION: PERIOSTIN-BASED DIAGNOSTIC ASSAYS
; FILE REFERENCE: 00530-099001
; CURRENT APPLICATION NUMBER: US/10/217,371
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US 60/312,123
; PRIOR FILING DATE: 2001-08-13
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 2250
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2250)
US-10-217-371-13

Query Match          63.8%; Score 16.6; DB 9; Length 2250;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      4 CTTGCAGATCGTACTCCTCAAT 26
         ||| ||| ||| ||| ||| ||| |||
DB      625 CTTACACAATTGTACTCCTCAAT 647

RESULT 9
US-10-217-371-3
; Sequence 3, Application US/10217371
; Publication No. US20030073137A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Ian Bo
; APPLICANT: Dai, Meiru
; APPLICANT: Sasaki, Hidefumi
; APPLICANT: Auciailt, Daniel
; TITLE OF INVENTION: PERIOSTIN-BASED DIAGNOSTIC ASSAYS
; FILE REFERENCE: 00530-099001
; CURRENT APPLICATION NUMBER: US/10/217,371
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US 60/312,123
; PRIOR FILING DATE: 2001-08-13
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2253
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2253)
US-10-217-371-3

Query Match          63.8%; Score 16.6; DB 9; Length 2253;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

OY 4 CTTTCGAGATCGTACCTCAAT 26
DB 688 CTTACACAATTTGTACTCTCAAT 710

RESULT 10
US-10-217-371-9

Sequence 9, Application US/10217371
Publication No. US20030073137A1

GENERAL INFORMATION:

APPLICANT: Chen, Ian Bo

APPLICANT: Dai, Meiru

APPLICANT: Sasaki, Hidefumi

APPLICANT: Auchi, Daniel

TITLE OF INVENTION: PERIOSTIN-BASED DIAGNOSTIC ASSAYS

FILE REFERENCE: 00530-099001

CURRENT APPLICATION NUMBER: US/10/217,371

PRIOR FILING DATE: 2002-08-13

PRIOR APPLICATION NUMBER: US 60/312,123

NUMBER OF SEQ ID NOS: 18

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 9

LENGTH: 2274

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (1)...(2274)

US-10-217-371-9

OY 4 CTTTCGAGATCGTACCTCAAT 26

DB 625 CTTACACAATTTGTACTCTCAAT 647

RESULT 11

US-10-217-371-11

Sequence 11, Application US/10217371

Publication No. US20030073137A1

GENERAL INFORMATION:

APPLICANT: Chen, Ian Bo

APPLICANT: Dai, Meiru

APPLICANT: Sasaki, Hidefumi

APPLICANT: Auchi, Daniel

TITLE OF INVENTION: PERIOSTIN-BASED DIAGNOSTIC ASSAYS

FILE REFERENCE: 00530-099001

CURRENT APPLICATION NUMBER: US/10/217,371

PRIOR FILING DATE: 2002-08-13

PRIOR APPLICATION NUMBER: US 60/312,123

NUMBER OF SEQ ID NOS: 18

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 11

LENGTH: 2313

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (1)...(2313)

US-10-217-371-11

OY 4 CTTTCGAGATCGTACCTCAAT 26

DB 688 CTTACACAATTTGTACTCTCAAT 710

Query Match 63.8%; Score 16.6; DB 9; Length 2313;

Best Local Similarity 82.6%; Pred. No. 1.1e+02;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

US-10-217-371-11

RESULT 12

US-10-217-371-7

Sequence 7, Application US/10217371

Publication No. US20030073137A1

GENERAL INFORMATION:

APPLICANT: Chen, Ian Bo

APPLICANT: Dai, Meiru

APPLICANT: Sasaki, Hidefumi

APPLICANT: Auchi, Daniel

TITLE OF INVENTION: PERIOSTIN-BASED DIAGNOSTIC ASSAYS

FILE REFERENCE: 00530-099001

CURRENT APPLICATION NUMBER: US/10/217,371

PRIOR FILING DATE: 2002-08-13

PRIOR APPLICATION NUMBER: US 60/312,123

NUMBER OF SEQ ID NOS: 18

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 7

LENGTH: 2337

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (1)...(2337)

US-10-217-371-7

OY 4 CTTTCGAGATCGTACCTCAAT 26

DB 688 CTTACACAATTTGTACTCTCAAT 710

RESULT 13

US-10-217-371-1

Sequence 1, Application US/10217371

Publication No. US20030073137A1

GENERAL INFORMATION:

APPLICANT: Chen, Ian Bo

APPLICANT: Dai, Meiru

APPLICANT: Sasaki, Hidefumi

APPLICANT: Auchi, Daniel

TITLE OF INVENTION: PERIOSTIN-BASED DIAGNOSTIC ASSAYS

FILE REFERENCE: 00530-099001

CURRENT APPLICATION NUMBER: US/10/217,371

PRIOR FILING DATE: 2002-08-13

PRIOR APPLICATION NUMBER: US 60/312,123

NUMBER OF SEQ ID NOS: 18

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1

LENGTH: 2508

TYPE: DNA

ORGANISM: Mus musculus

FEATURE:

NAME/KEY: CDS

LOCATION: (1)...(2508)

US-10-217-371-1

OY 4 CTTTCGAGATCGTACCTCAAT 26

DB 688 CTTACACAATTTGTACTCTCAAT 710

Query Match 63.8%; Score 16.6; DB 9; Length 2508;

Best Local Similarity 82.6%; Pred. No. 1.1e+02;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

US-10-217-371-1

```

US-10-176-847-45
: Sequence 45, Application US/10176847
: Publication No. US20030068636A1
: GENERAL INFORMATION:
: APPLICANT: Velby, Pelter Ole
: TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
: TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
: TITLE OF INVENTION: AND OVARIAN CANCER
: FILE REFERENCE: MRI -039
: CURRENT APPLICATION NUMBER: US/10/176,847
: CURRENT FILING DATE: 2002-06-21
: NUMBER OF SEQ ID NOS: 112
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 45
: LENGTH: 2937
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-10-176-847-45

```

Query Match	63.8%	Score 16.6	DB 9	Length 2937
Best local similarity	82.6%	Pred. No. 1.1e+02		
Matches 19; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

Oy	4	CTTTCGAGATCGGTACCTCAAT	26
Db	768	CTTACACAATTTGGTACCTCAAT	790

```

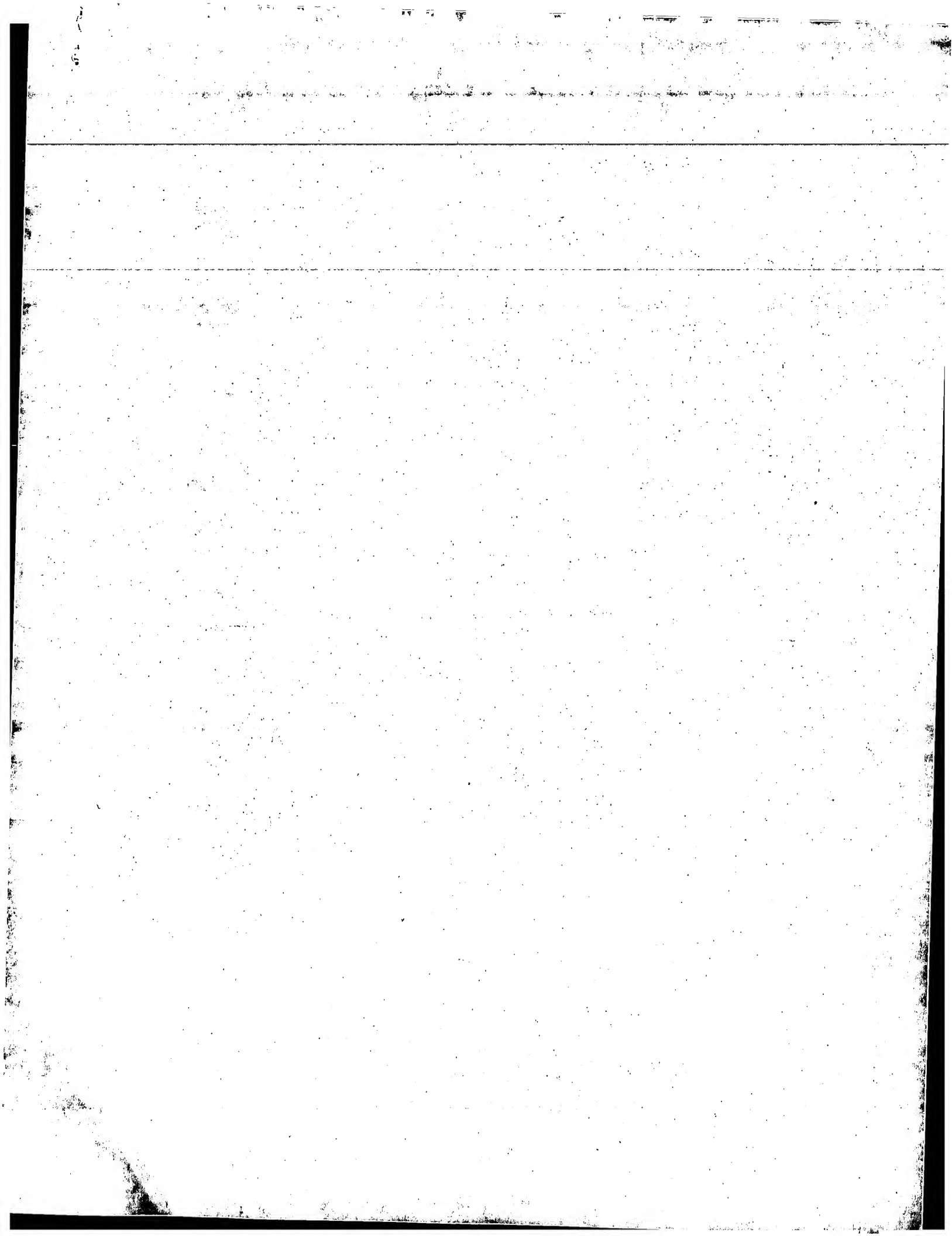
RESULT 15
US-10-171-311-179
: Sequence 179, Application US/10171311
: Publication No. US20030087270A1
:
GENERAL INFORMATION:
:
APPLICANT: Schlegel, Robert
:
APPLICANT: Chen, Yan
:
APPLICANT: Zhao, Xumei
:
APPLICANT: Monahan, John
:
APPLICANT: Kamatkar, Shubhangi
:
APPLICANT: Glatf, Karen
:
APPLICANT: Gannavarapu, Manjula
:
APPLICANT: Hoersch, Sebastian
:
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
:
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
:
TITLE OF INVENTION: OF CERVICAL CANCER
:
FILE REFERENCE: MRI-035
:
CURRENT APPLICATION NUMBER: US/10/171,311
:
CURRENT FILING DATE: 2002-06-12
:
PRIOR APPLICATION NUMBER: US 60/298,159
:
PRIOR FILING DATE: 2001-06-13
:
PRIOR APPLICATION NUMBER: US 60/298,155
:
PRIOR FILING DATE: 2001-06-13
:
PRIOR APPLICATION NUMBER: US 60/335,936
:
PRIOR FILING DATE: 2001-11-14
:
NUMBER OF SEQ ID NOS: 238
:
SOFTWARE: FastSeq for Windows Version 4.0
:
SEQ ID NO 179
:
LENGTH: 3077
:
TYPE: DNA
:
ORGANISM: Homo sapiens
:
US-10-171-311-179

```

Query Match	63.88	Score	16.6	DB	9	Length	3077
Similarity	82.68	Pred. No.	1.1e+02				
Best Local							
Matches	19	Conservative	0	Mismatches	4	Indels	0
						Gaps	0

OY	4	CTTTCGAGATCGGTACCTCAAT	26
Db	715	CTTACACAATTTGGTACCTCAAT	737

Search completed: July 11, 2003, 15:02:06
Job time : 80.164 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:13:45 ; Search time 467.474 Seconds
(without alignments)
900.761 Million cell updates/sec

Title: US-10-087-631B-8

Sequence: 1 cgtcttcgcagctgactcaat 26

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmuv:*
5: em_estov:*
6: em_estipl:*
7: em_estiro:*
8: em_estc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.6	75.4	961	17	AO900643 HS_3171_A
2	19.2	73.8	185	10	BB600244 BB600244
3	19.2	73.8	698	13	BI924556 EST544445
4	18.6	71.5	460	9	AI351632 q05h12.x
5	18.6	71.5	1029	12	BF098973 601750475
6	18.2	70.0	182	17	CNS02NMJM AL205483 Tetraodon

Result No.	Score	Query Match	Length	DB ID	Description
7	18.2	70.0	234	10	BB593576 BB593576
8	18.2	70.0	242	10	BB600933 BB600933
9	18.2	70.0	243	10	BB569672 BB569672
10	18.2	70.0	290	10	BB592646 BB592646
11	18.2	69.2	312	10	BB252299 BB252299
12	18.2	69.2	411	17	BB588219 BB588219
13	18.2	69.2	435	12	BF588219 BF588219
14	18.2	69.2	701	17	AO255220 AO255220
15	17.6	67.7	150	12	BF757321 BF757321
16	17.6	67.7	285	13	BI172474 BI172474
17	17.6	67.7	335	10	AV244463 AV244463
18	17.6	67.7	381	9	AA538976 AA538976
19	17.6	67.7	386	14	BM870563 BM870563
20	17.6	67.7	421	17	AI247865 AI247865
21	17.6	67.7	454	13	BI164958 BI164958
22	17.6	67.7	498	13	BI171702 BI171702
23	17.6	67.7	512	17	BI177119 BI177119
24	17.6	67.7	512	17	CNS07J0J CNS07J0J
25	17.6	67.7	527	13	BI227641 BI227641
26	17.6	67.7	547	13	BI170103 BI170103
27	17.6	67.7	557	13	BI362843 BI362843
28	17.6	67.7	578	9	AA263759 AA263759
29	17.6	67.7	578	10	AV517896 AV517896
30	17.6	67.7	581	13	BI213740 BI213740
31	17.6	67.7	589	13	BI168983 BI168983
32	17.6	67.7	590	14	BI6252 BI6252
33	17.6	67.7	594	9	AI515767 AI515767
34	17.6	67.7	594	9	AI999483 AI999483
35	17.6	67.7	599	13	BI169019 BI169019
36	17.6	67.7	605	13	BI171671 BI171671
37	17.6	67.7	618	13	BI356268 BI356268
38	17.6	67.7	618	9	AA820697 AA820697
39	17.6	67.7	624	13	BI485494 BI485494
40	17.6	67.7	627	13	BI230919 BI230919
41	17.6	67.7	642	13	BI169762 BI169762
42	17.6	67.7	659	13	BI238807 BI238807
43	17.6	67.7	659	17	BI175201 BI175201
44	17.6	67.7	659	17	CNS0718R CNS0718R
45	17.6	67.7	662	17	BB655017 BB655017

ALIGNMENTS

RESULT 1
AO900643
LOCUS
DEFINITION
HS_3171_A2_A07_TTC CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone plate-3171 Col-14 Row-A, DNA sequence.
ACCESSION
AO900643
VERSION
AO900643.1 GI:6356833
KEYWORDS
GSS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 961)
Mahalax, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
CONTACT: Mahalax GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: <http://www.htsc.washington.edu>

Query Match	75.48;	Score 19.6;	DB 17;	Length 961;
Best Local Similarity	84.68;	Pred. No. 1.5e+02;		
Matches 22;	Conservative	0;	Mismatches 4;	Indels 0;
				Gaps 0

Qy	1	CGTCCTTCGCAGATCGGTACCTCAAT	26
Db	711	CGTCGTGCGGGTTCGGTACCAACAAT	731

RESULT 2	BB600244/c				
LOCUS	BB600244	185 bp	mRNA	linear	EST 01-DEC-2000
DEFINITION	BB600244	RIKEN full-length enriched, 12 days embryo	spinal ganglion		
ACCESSION	BB600244	Mus musculus cDNA clone D130004I24.5,	mRNA sequence.		
VERSION	BB600244.1	GI:11508485			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				

REFERENCE	AUTHORS
1 (bases 1 to 185)	Alizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, F., Hanagaki, T., Hayatsu, N., Hiraoka, T., Hirozane, T., Hodyama, Y., Imotani, K., Ishii, Y., Itoh, M., Iwawa, M., Kawai, J., Kojima, Y., Komura, H., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owe, C., Sakai, C., Saito, K., Saeki, D., Seto, K., Shibata, K., Shihata, Y., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka, T., Toyota, T., Watanabe, A., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshiki, A., Muramatsu, M., and Hayashizaki, Y.
TITLE	RIKEN Mouse ESTs (Alizawa, K. et al. 2000)
JOURNAL	Unpublished (2000)
COMMENT	Contact: Yoshinhide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9226
Email: genome-res@gs.c.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagoka, S., Sasakawa, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermotranscription and thermotranscription of thermolabile enzymes by
thermolabile and its application for the synthesis of full length
cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)
Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. *Genome Res.* 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303,
19-44 (1999)
Please visit our web site (<http://genome.ri.c.riken.go.jp>) for
further details.

FEATURES:
Source

```

Location/Qualifiers
1. 185
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="ID130004124"
/clone_idb="Riken full-length enriched, 12 days embryo
spinal ganglion"
/tissue-type="spinal ganglion"
/dev_stage="12 days embryo"
/lab_host="DH108"

```

BASE COUNT
ORIGIN

42 a 44 c 61 g 38 t

/mode=site.1; salt; site.2; BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken, Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15' GAGGAGAGAGCGCGCCGACATCGAGTTTTTTTTTTTTTNN 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGGAGAGATTCGAGTTAAATTAATATATCCCGCCCCCCC 3'. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda E1C I."

Query Match	73.88;	Score 19.2;	DB 10;	Length 185;
Best Local Similarity	87.5%;	Pred. No. 1.7e+02;		
Matches 21; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

QY 3-TCTTTGCAGATCGGTACTTCAAT 26
 Db . ||||| | | ||||| |||
 48 TCTTTCGAATCGGTACTTCAAT 25

[illegible]

REFERENCE AUTHORS

FEATURES	Location/Qualifiers
source	1. 698

```
Location/Qualifiers
1. 698
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cROA239"
/clone_1lb="tomato flower, buds 0-3 mm"
/tissue_type="flower"
```


/dev_stage="0-3mm buds"
 /note="Vector: Bluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; supplier: Cornell University; sequencing: The
 Institute for Genomic Research; Flower buds and flowers
 were taken from greenhouse plants (4-8 wks old, T496).
 They were immediately frozen in liquid nitrogen and then
 size-separated while remaining frozen."
 BASE COUNT 214 a 120 c 166 g 198 t
 ORIGIN

Query Match 73.8%; Score 19.2; DB 13; Length 698;
 Best Local Similarity 87.5%; Pred. No. 2.2e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 3 TCTTCCGACATCGTACTCAAT 26
 Db 133 TTTTCTCGATGGTACTCAAT 110

RESULT 4
 LOCUS AT151632 460 bp mRNA linear EST 13-FEB-1999
 DEFINITION qf05h12.x1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
 IMAGE:1940039 3', mRNA sequence.
 ACCESSION AT151632
 VERSION AT151632.1 GI:4088838
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 460)
 AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (linfoimage.lnl.gov) for further information.
 Insert Length: 1001 Std Error: 0.00
 Seq primer: -400p from Gldco
 High quality sequence stop: 437.
 FEATURES
 location/Qualifiers
 source 1..460
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:1940039"
 /clone_lib="Soares_total_fetus_Nb2HF8_9w"
 /dev_stage="8-9 weeks"
 /lab_host="DH10B"
 /note="Vector: pT733-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from pooled 8-9 week
 (total) fetus material with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAAGGAGCGCGCTTAATTTTCTTTTCTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT733 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 131 a 91 c 107 g 131 t
 ORIGIN

Query Match 71.5%; Score 18.6; DB 9; Length 460;
 Best Local Similarity 84.0%; Pred. No. 3.7e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 GTCTTCCGACATCGTACTCAAT 26
 Db 422 GTCTTCCGACATCGTACTCAAT 398

RESULT 5
 LOCUS BF098973/c 1029 bp mRNA linear EST 19-OCT-2000
 DEFINITION 601750475p1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3978234 5',
 mRNA sequence.
 ACCESSION BF098973
 VERSION BF098973.1 GI:10881499
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
 REFERENCE 1 (bases 1 to 1029)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LHM9170 row: e column: 19
 High quality sequence stop: 636.
 FEATURES
 location/Qualifiers
 source 1...1029
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_image="IMAGE:3978234"
 /clone_lib="NCI_CGAP_Mam1"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="10 months, virgin"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT 257 a 295 c 300 g 177 t
 ORIGIN

Query Match 71.5%; Score 18.6; DB 12; Length 1029;
 Best Local Similarity 84.0%; Pred. No. 4.4e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 GTCTTCCGACATCGTACTCAAT 26
 Db 825 GTCTTCCGACATCGTACTCAAT 801

RESULT 6
 LOCUS CNS02NM/c 182 bp DNA linear GSS 14-MAY-2000
 DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
 151119 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL205483
 VERSION AL205483.1 GI:7864302
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 182)
 Roest-Crollius, R., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fitzames, C., Wincker, P., Brothier, P., Queller, F.,
 Saurin, W. and Weissbach, J.
 Human gene number estimate provided by genome wide analysis using

[illegible]

ACCESSION	BB569672
VERSION	BB569672.1
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

Query Match	70.0%	Score 18.2	DB 10	Length 243
Best Local Similarity	87.0%	Pred. No. 5e+02		
Matches 20	Conservative	0	Mismatches 3	Indels 0
				Gaps 0
4	CTTTCGCAGATCGGTACTCTCAAT	26		

32 CTTCTCAATCGGTACCTTCAT 10

LOCUS	290 bp	mRNA	linear	EST	30-NOV-2000
BB592646					
BB592646	RIKEN	full-length enriched, adult male	cornu		

VERSION BB592646.1 GI:11489248
KEYWORDS EST

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus

'Hanagaki, T., Hayatsu, N., Hiraoka, T., Hirozane, T., Hodozama, Y.,
Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, I., Kotima, Y., Kono,

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resesc.riken.go.jp,
URL: <http://genome-gsc.riken.go.jp/>
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermosensitization and thermoinactivation of thermolabile enzymes by
carnalose and its application for the synthesis of full length
cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)
Itoh, M., Kitanai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
Y. and Hayashizaki, Y.,
Automated filtration-based high-throughput plasmid preparation
system. *Genome Res.* 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303,
19-44 (1999)
Please visit our web site (<http://genome.riken.go.jp/>) for
further details.

```

/organism="Mus musculus"
/db_xref="taxon:10090"

```

```

/cclone="A230104022"
/cclone_1lb="Riken full-length enriched, adult male corpora
quadrigemina"
/sex="male"
/tissue_type="corpora quadrigemina"
/dev_stage="adult"
/lab_host="DHI0B"
/note="Site-1: Salt; Site-2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGGAGAGGAGACGCCAGACACTCTTTTATTTTTTTTTTNN 3']. cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by

```

Query Match	70.0%	Score 18.2	DB 10	Length 290
Best Local Similarity	87.0%	Pred. No. 5, 2e+02		
Matches 20	Conservative	0	Mismatches 3	Indels 0
				Gaps 0

DROITE 19

RESULT 11					
LOCUS					
BB252299					
DEFINITION	BB252299	312 bp	mRNA	linear	EST 06-JUL-2000
	BB252299	RIKEN	full-length enriched,	7 days neonate	cerebellum Mms

Accession BB252299

KEYWORDS	EST.
SOURCE	house mouse.

REFERENCE
AUTHORS

Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arikawa, T., Carolin, I. (bases 1 to 312)

Mammalia; Eutheria; Rodentia; Caniata; Vertebrata; Euteleostomi; Muriinae; Mus.

Sciurognathi; Muridae; Muriinae; Mus.

P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, U., Kikuchi, N., Kiyosawa, H., Kojima, T., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomiaga, N., Toyota, T., Tsunoda, Y., Wataniki, A., Watanabe, S., Yamamura, T., Yamana, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.

RIKEN MOUSE ESTS (Konno, H., et al.)

Unpublished (2000)

TITLE
JOURNAL
COMMENT

Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 225-8505, Japan

Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URI: <http://genome.gsc.riken.go.jp/>,
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermotabilization and thermocycling of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

1. .312
location/Qualifiers

FEATURES	source
BASE COUNT	63 a 178 c 107 g 87 t
ORIGIN	
Query Match	69.2%; Score 18; DB 12; Length 435;
Best Local Similarity	80.8%; Pred. No. 6.8e+02;
Matches	21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
1 CGTCTTGCAGATCGGTACTCAAT 26	
219 CGTCTTGCAGATCGGTACTCAAT 244	
BASE COUNT	98 a 108 c 100 g 103 t 2 others
ORIGIN	
Query Match	69.2%; Score 18; DB 17; Length 411;
Best Local Similarity	80.8%; Pred. No. 6.8e+02;
Matches	21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
1 CGTCTTGCAGATCGGTACTCAAT 26	
57 CGTCTTGCAGATCGGTACTCAAT 32	
BASE COUNT	63 a 178 c 107 g 87 t
ORIGIN	
Query Match	69.2%; Score 18; DB 12; Length 435;
Best Local Similarity	80.8%; Pred. No. 6.8e+02;
Matches	21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
1 CGTCTTGCAGATCGGTACTCAAT 26	
219 CGTCTTGCAGATCGGTACTCAAT 244	

RESULT 14
LOCUS A0255220
DEFINITION mgx0009D16r CUG1 Rice Blast BAC Library Magnaporthe grisea genomic clone mgx0009D16r, DNA sequence.
ACCESSION A0255220
VERSION A0255220.1 GI:3779535
KEYWORDS GSS.
SOURCE Magnaporthe grisea.
ORGANISM Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE 1 (bases 1 to 701)
YU, Y., ZHU, H., BOYD, C.A., GAUDETTE, B., GAYLE, A., KINGSBURY, R., PHILLIPS, R., SASNOWSKI, M., WING, R.A. and DEAN, R.A.
A BAC End Sequencing Framework to Sequence the Magnaporthe grisea Genome
TITLE Unpublished (1998)
JOURNAL Contact: Dean RA
COMMENT Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdeane@clemson.edu
Seq primer: GGAAACACCTATGACCATG
Class: BAC ends
High quality sequence stop: 445.
Location/Qualifiers
1..701
/organism="Magnaporthe grisea"
/strain="70-15"
/db_xref="taxon:148305"
/clone="mgx0009D16r"
/clone_lib="CUG1 Rice Blast BAC Library"
/tissue_type="Protoplasts"
/lab_host="E. coli DH10B"
/note="Vector: pBACWICH; Site_1: HindIII; Site_2: HindIII; Rice blast is one of the most devastating fungal diseases of rice world wide. It is a filamentous ascomycete with a haploid genome (n=7) of approximately 40 Mbp. Rice blast is an important model fungal pathogen for studying numerous aspects of the fungal-host interaction. In order to facilitate genome wide analysis, a BAC library containing 19216 clones with an average insert size of 130 kbp was constructed. This library represents greater than 25x genome coverage. High density colony filters are available upon request."
BASE COUNT 168 a 156 c 170 g 206 t 1 others
ORIGIN
Query Match 69.2%; Score 18; DB 17; Length 701;
Best Local Similarity 100.0%; Pred. No. 7.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 GCAGATCGTACCTCAAT 26
DB 405 GCAGATCGTACCTCAAT 422
RESULT 15
LOCUS BF757321
DEFINITION MRO-CT0452-041100-303-h10 CT0452 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF757321
VERSION BF757321.1 GI:12105221
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 150)

AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W.-Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
TITLE Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL Contact: Simpson A.J.G.
COMMENT Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MROct2-MRO-CT0452-041100-303-h10&t3=2000-11-04&t4=1)
Seq primer: puc 18-forward
High quality sequence start: 13
High quality sequence stop: 150.
Location/Qualifiers
1..150
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0452"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,776 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 32 a 40 c 46 g 32 t
ORIGIN
Query Match 67.7%; Score 17.6; DB 12; Length 150;
Best Local Similarity 83.3%; Pred. No. 8.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 GTCTTCGACATCGACCTCAA 25
DB 89 GTCTTCGACATCGACCTCAA 66

Search completed: July 11, 2003, 02:26:05
Job time : 474.474 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 242.625 Seconds

(without alignments)
3358.597 Million cell updates/sec

Title: US-10-087-631B-9

Perfect score: 28

Sequence: 1 cgttcgtgggatacgtcgtcatgtgtt 28

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb_da:*

2: gb_htg:*

3: gb_in:*

4: gb_cm:*

5: gb_ov:*

6: gb_ph:*

7: gb_pl:*

8: gb_pr:*

9: gb_ro:*

10: gb_sts:*

11: gb_sy:*

12: gb_un:*

13: gb_vl:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_mu:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rtd:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20.2	72.1	193862	2	AC125937
2	20	71.4	10100	1	AE005060
3	19.6	70.0	187487	2	AC119388
4	19.6	70.0	191844	2	AC122410
5	19.2	68.6	34388	3	CER2888
6	19	67.9	1841	3	DDICP100
7	19	67.9	11884	1	AE004745
8	19	67.9	14687	2	AC126800
9	19	67.9	80015	9	AC110085
10	19	67.9	115045	9	AC004414
11	19	67.9	123770	2	AC120560
12	19	67.9	153987	2	AC094593
13	19	67.9	179572	2	AC117647
14	19	67.9	189086	2	AC009921
15	19	67.9	209743	2	AL354896
16	19	67.9	276248	2	AC087561
17	18.8	67.1	120033	2	AC122725
18	18.6	66.4	90052	2	AL139429
19	18.6	66.4	110120	2	AC117287
20	18.6	66.4	129296	9	AC104123
21	18.6	66.4	129667	2	AC095300
22	18.6	66.4	143783	2	AC119005
23	18.6	66.4	145514	9	AC095509
24	18.6	66.4	154699	2	AC115657
25	18.6	66.4	174342	2	AC120962
26	18.6	66.4	184061	2	AC121189
27	18.6	66.4	190212	2	AC127097
28	18.6	66.4	191597	2	AC121956
29	18.6	66.4	191992	2	AC097815
30	18.6	66.4	233488	2	AC122228
31	18.6	66.4	240050	1	AL627267
32	18.4	65.7	795	3	AF291442
33	18.4	65.7	801	3	AF291441
34	18.4	65.7	2578	3	AY051922
35	18.4	65.7	4218	1	AF262949
36	18.4	65.7	66497	2	AC015388
37	18.4	65.7	66497	2	AC016426
38	18.4	65.7	97146	8	AC007259
39	18.4	65.7	110773	9	AC012601
40	18.4	65.7	141867	9	AL450445
41	18.4	65.7	149980	2	AP003612
42	18.4	65.7	151386	2	AC013312
43	18.4	65.7	151386	2	AC013312
44	18.4	65.7	158174	2	AC123065
45	18.4	65.7	158923	9	AC097377

ALIGNMENTS

RESULT 1

AC125937

LOCUS: Rattus norvegicus clone CH230-11B2, *** SEQUENCING IN PROGRESS ***

DEFINITION: 57 unordered pieces.

ACCESSION: AC125937.1 GI:21671543

VERSION: HTG: HTGS_PHASE1.

KEYWORDS: Norway rat.

SOURCE: Rattus norvegicus

ORGANISM: Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 193862)

AUTHORS: Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Ayale, M., Banks, T.,
 Barbara, J., Benton, J., Blinige, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowle, S., Brileva, M., Brown, M., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Butrell, K.L., Byrd, N.C.,
 Caron, T.E., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Galis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J., Harris, K., Hart, M., Havlak, P., Hale, S., Hamilton, K.,
 Hernandez, O., Hodgson, A., Hogue, M., Hollway, C., Hollins, B.,
 Homsl, F., Howard, S., Huber, J., Huylk, S., Kume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karason, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratoch, J., Kuesel, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichte, O., Lieu, C., Liu, J., Liu, W., Lounseged, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawlin, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, U., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okunolu, G.,
 Otunuga, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoshart, N., Sisson, H.,
 Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansy, U., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalobos, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

Unpublished
 Direct Submission
 2 (bases 1 to 193862)
 Worley, K.C.

Submitted (02-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 193862)
 Worley, K.C.

Direct Submission
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: GDM
 Center clone name: CH230-1182
 Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 143066 bases at least Q40
 Consensus quality: 151194 bases at least Q30
 Consensus quality: 158097 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 57 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	1224	contig of 1224 bp in length
1	1325	gap of unknown length
1	1325	contig of 1257 bp in length
1	2582	gap of unknown length
1	2681	contig of 1677 bp in length
1	4358	gap of unknown length
1	4459	contig of 1545 bp in length
1	6004	gap of unknown length
1	6104	contig of 1642 bp in length
1	7745	gap of unknown length
1	7746	contig of 1380 bp in length
1	7846	gap of unknown length
1	9226	contig of 1024 bp in length
1	9326	gap of unknown length
1	10350	contig of 1809 bp in length
1	10450	gap of unknown length
1	12259	contig of 1283 bp in length
1	12359	gap of unknown length
1	13642	contig of 1128 bp in length
1	13741	gap of unknown length
1	14869	contig of 1128 bp in length
1	14970	gap of unknown length
1	16977	contig of 2007 bp in length
1	17077	gap of unknown length
1	17077	contig of 1707 bp in length
1	18784	gap of unknown length
1	18884	contig of 2015 bp in length
1	20899	gap of unknown length
1	20999	contig of 1891 bp in length
1	22890	gap of unknown length
1	22989	contig of 1838 bp in length
1	24827	gap of unknown length
1	24927	contig of 1066 bp in length
1	25993	gap of unknown length
1	26093	contig of 1731 bp in length
1	27825	gap of unknown length
1	27924	contig of 1897 bp in length
1	29821	gap of unknown length
1	29921	contig of 2101 bp in length
1	32022	gap of unknown length
1	32122	contig of 1966 bp in length
1	32123	gap of unknown length
1	34089	contig of 1936 bp in length
1	34189	gap of unknown length
1	36124	contig of 1936 bp in length
1	36224	gap of unknown length
1	37933	contig of 1709 bp in length
1	38033	gap of unknown length
1	38034	contig of 1522 bp in length
1	39555	gap of unknown length
1	39656	contig of 2063 bp in length
1	41719	gap of unknown length
1	41818	contig of 1663 bp in length
1	43481	gap of unknown length
1	43482	contig of 1663 bp in length
1	43582	gap of unknown length
1	46430	contig of 2849 bp in length
1	46431	gap of unknown length
1	46531	contig of 2517 bp in length
1	49047	gap of unknown length
1	49147	contig of 2027 bp in length
1	51174	gap of unknown length
1	51274	contig of 4570 bp in length
1	55844	gap of unknown length
1	55945	contig of 2293 bp in length
1	55945	gap of unknown length
1	58237	contig of 2570 bp in length
1	58338	gap of unknown length
1	60908	contig of 3716 bp in length
1	61007	gap of unknown length
1	64724	contig of 3110 bp in length
1	64823	gap of unknown length
1	67933	contig of 3110 bp in length
1	68033	gap of unknown length
1	70343	contig of 2310 bp in length

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)

JOURNAL MEDLINE 20504483

REFERENCE PubMed 11016950

AUTHORS 2 (bases 1 to 10100)

1. W.V., Kennedy, S.P., Mahairas, G.G., Bergquist, B., Pan, M., Shukla, H.D., Lasky, S.R., Baliga, N., Thorsson, V., Shrogha, J., Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Melli, R., Goo, Y.A., Leitbauer, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W., Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Dale, H., Isenbarger, T.A., Peck, R.F., Fohnschrod, M., Spadich, U.L., Jung, K.-H., Alam, T., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P., Omer, A.D., Ehardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L. and Dassarma, S.

Direct Submission

TITLE Submitted (14-JUL-2000) Institute for Systems Biology, 4225

JOURNAL Roosevelt Way NE, Seattle, WA 98105, USA

FEATURES

Source Location/Qualifiers

1. 10100

/organism="Halobacterium sp. NRC-1"

/strain="NRC-1"

/db_xref="taxon:64091"

115. 570

/gene="VNG1420H"

115. 570

/gene="VNG1420H"

/note="hypothetical protein"

/codon_start=1

/transl_table=1

/product="Vng1420h"

/protein_id="AAG19736.1"

/db_xref="GI:10580924"

/translation="MRLVEKFPAGSPDCAWRGRGDRVVVSVKGDELVLPTDAEQLRDS
LSPALTRRESFMTAGVRPDSYVWRKADSGNKRVESEFALGRCRRRPETFTT
AEULSATGAGRRHVMVLVEHPAPVALASNPPLTAETAREVAEP"

601. 1512

/gene="VNG1422H"

601. 1512

/gene="VNG1422H"

/note="hypothetical protein"

/codon_start=1

/transl_table=1

/product="Vng1422h"

/protein_id="AAG19737.1"

/db_xref="GI:10580925"

/translation="MRFSTVARDEPTDAVAVQSKFVAVGSPFASADAGAIATQSS
LAWVAGSDLDLIRAGHSADAEVAELTNTDDEAPQKQVGVGGDSVAFTGAECVVD
HASDRGDHTTAQGNILANRATVADAMADTFEDFGGLPERLLALHAGNDAGCGQREB
OGALIVYAKPGGSGDGNIDIVRVDDHHPIDELRRVHIYDITLLERVAPEYEEK
IADDTAEVTSALSGISGFGYDCAIPGRARGDADRDLAAFOAVNNENRNDLALAEALAR
GNADAGESEDRLYNATWQGISRLDR"

1571. 1717

/gene="VNG1423H"

1571. 1717

/gene="VNG1423H"

/note="hypothetical protein"

/codon_start=1

/transl_table=1

/product="Vng1423h"

/protein_id="AAG19738.1"

/db_xref="GI:10580926"

/translation="MRDEEIRIQEYFLVEELSEEMSHGIREMFTYRRLGMALE
EEMH"

complement(1928. 2224)

/gene="VNG1425H"

complement(1928. 2224)

/gene="VNG1425H"

/note="hypothetical protein"

/codon_start=1

/transl_table=1

/product="Vng1425h"

/protein_id="AAG19739.1"

/db_xref="GI:10580927"

/translation="MTEFTVFESSTQSRADVAAYLTVAIDNQNNDITILSAGDPE

```

gene
2365. .2643
/gene="VNG1426H"
2365. .2643
/gene="VNG1426H"
/note="hypothetical protein"
/codon_start=1
/product="Vng1426h"
/protein_id="AAG19740.1"
/db_xref="GI:10580928"
/translation="MMDLTFQFDLLVYVAGLDEPHGLAKDELQYLYSEIHGRLY
PMLDTLVEKGLDKRGELDQRTNYTLTRGTRLEAREMRETOYVAGS"
2647. .2889
/gene="VNG1427H"
2647. .2889
/gene="VNG1427H"
/note="hypothetical protein"
/codon_start=1
/product="Vng1427h"
/protein_id="AAG19741.1"
/db_xref="GI:10580929"
/translation="MYARKLATAVAGVAGSLAVSVVAMQVFDLPVFLVAVPEPFLR
KRGSSPRTKRCPECGFTYRTPSVTHCHDGTELD"
2647. .4034
/gene="hclA"
2647. .4034
/complement(2886. .4034)
/note="hclA"
/codon_start=1
/product="Hcl-like protein"
/protein_id="AAG19742.1"
/db_xref="GI:10580930"
/translation="MGDTRIDRGRAAMALIGTTLAALVYVYVGTGTVLGFLLYA
TRPVRRVDRHRSSVAAYVSLTALPVVLSYTTAVGIDELTRVLPQIDLT
AVLAVQPYDVSRANQDQSLGQPSVAAINTLRSAGVAGFLGTAFLFAVIT
FNLQIVPSSLSIPYPIKMLTLCGASLIPVGIKVIWILLGLMVAARSGSGG
FVAVGVAAVAVTIFIDLVVRYVYSGDLHGLMVLAVYVGPLTWYIGTFLCPML
VAVVHFAVLVPELVAGGIAVPAGADDTASADPESADPDDTA"
4126. .5136
/gene="VNG1429C"
4126. .5136
/gene="VNG1429C"
/note="conserved hypothetical protein"
/codon_start=1
/translation="MTFLAGGTGTPRLLAGARRVDPATTVVNTGDDVALGLLY
CPDLDVLEGGVLDRETWGLADDSSTHDTYTLDAALADIDPDPRLPDDATA
GRDLARRRSSASEMFGIDRBAVHTLRAGLIDEGHTLETYRLRLADFDLSDLY
PMSNDPVAIVOTPDGEQHOTFVVAHGDPTVEDVEFGEGEATAPRATERRICP
VAVGSPNPTSIGPMLADIGLADLRDAOVAVASPRVEDVEFGPAKMAAAGHDS
TAGVADAVPADAFVLDTASTDLDLRVYVTDLSLDEADAEVAVARCRDALVAASGE
VT"
5198. .5869
/gene="VNG1431C"
5198. .5869
/gene="VNG1431C"
/note="conserved hypothetical protein"
/codon_start=1
/translation="Vng1431c"
/protein_id="AAG19744.1"
/db_xref="GI:10580932"
/translation="MSAGEHAGGVFLGIAIDEPREARARAVDNDREFFLPADVAV
VDDQLSELDALFRAVYVAVRATGLRLRYAAVCAHDLITLVNAHCRQDELCDAGAG
ETLRDTSRCGEVRAASDEGAASVAVKAVAGVGVHLPLVLAQATERRAGAAAIHVAMD

```

```

gene
6000. .6992
/gene="dhs"
6000. .6992
/note="VNG1432C"
6000. .6992
/gene="dhs"
/note="Dhs"
/codon_start=1
/translation="deoxyhypusine synthase"
/protein_id="AAG19745.1"
/db_xref="GI:10580933"
/translation="MTGDDAEFHENNVPSDEDLDLPDVRGNDPSCGFPEFLDST
ATGQASHADAVDITPEMEREDATYITLITSIVSSGLREVVAHLVRENYDVIT
TSGSTEDITITKARPKRGENDVDEALREKGINRLNIPVSDRYVWLEBYLDFE
EPFADQKRTPTAFARRELATDEDSITLKNADNDIPVCPALTDALGNFLYVRO
GYDEVGTEIILDYDALIEGLDADTGLICVGVGPKHAIIMTLNFERGADVAVYIS
TGMGDSLSGAPPEEAVSMCKIRDEDAEPYALIEATLVPLVAGAFENP"
7078. .7257
/gene="rps17e"
7078. .7257
/note="VNG1433C"
7078. .7257
/gene="rps17e"
/note="Rps17e"
/codon_start=1
/translation="30S ribosomal protein S17e"
/protein_id="AAG19746.1"
/db_xref="GI:10580934"
/translation="MALKPDYVKKTGNIIMERYODAFSRDEPHNKDAVTELTNIESK
VNRKRIAGYTHRN"
7304. .8335
/gene="asd"

Query Match 71.4% Score 20; DB 1; Length 10100;
Best Local Similarity 82.1% Pred. No. 51;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGTTCTGGGATATCCGTCATGCTGT 28
Db 3946 CGTTCTGGGATATCCGTCATGCTGT 3919

RESULT 3
AC119388
LOCUS
DEFINITION
Rattus norvegicus clone CH230-215C17, *** SEQUENCING IN PROGRESS
ACCESSION
AC119388.6 GI:21902577
VERSION
HTG: HTGS_PHASE1.
KEYWORDS
Rattus norvegicus.
SOURCE
Rattus norvegicus.
ORGANISM
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 187487)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbarta,U., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Boyle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dethorne,S.R., David,R.,
Devilay,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,

```

Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homi,F., Howard,S., Huber,J., Huliyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kravic,U., Kuresh,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Meador,M.P., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuon,G., Oregany,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Plickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rotubokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,O., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and Gibbs,R.

Unpublished
2 (bases 1 to 187487)
Worley,K.C.
Direct Submission
Submitted (26-Apr-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 187487)
Worley,K.C.
Direct Submission
Submitted (23-Jul-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 18, 2002 this sequence version replaced gi:20467907.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWKF
Center clone name: CH230-215C17
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 14644 bases at least Q40
Consensus quality: 153280 bases at least Q30
Consensus quality: 157543 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draift_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 54 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1224 1223: contig of 1223 bp in length
* 1324 1223: gap of unknown length
* 1324 2465: contig of 1142 bp in length
* 2465 2565: gap of unknown length
* 2565 3894: contig of 1329 bp in length
* 3894 3895 3895: gap of unknown length
* 3895 5712: contig of 1718 bp in length
* 5712 5712: gap of unknown length

5713 5812: gap of unknown length
5813 6812: contig of 1000 bp in length
6813 8166: gap of unknown length
8166 8613 8613: contig of 1254 bp in length
8613 8266: gap of unknown length
8266 9401: contig of 1135 bp in length
9401 9502 9502: gap of unknown length
9502 10959: contig of 1458 bp in length
10959 11059: gap of unknown length
11059 12324: contig of 1265 bp in length
12324 12424: gap of unknown length
12424 14011: contig of 1587 bp in length
14011 14111: gap of unknown length
14111 15777: contig of 1666 bp in length
15777 15877: gap of unknown length
15877 17879: contig of 2002 bp in length
17879 17880 17880: gap of unknown length
17880 17979: gap of unknown length
17979 19661: contig of 1662 bp in length
19661 19761: gap of unknown length
19761 21649: contig of 1888 bp in length
21649 21749: gap of unknown length
21749 23612: contig of 1863 bp in length
23612 23712: gap of unknown length
23712 25660: contig of 1948 bp in length
25660 25760: gap of unknown length
25760 27188: contig of 1428 bp in length
27188 27289: gap of unknown length
27289 29615: contig of 2327 bp in length
29615 29715: gap of unknown length
29715 32390: contig of 2675 bp in length
32390 32490: gap of unknown length
32490 33491 33491: contig of 2562 bp in length
33491 35052: contig of 2562 bp in length
35052 35152: gap of unknown length
35152 37301: contig of 2149 bp in length
37301 37401: gap of unknown length
37401 39449: contig of 2048 bp in length
39449 39549: gap of unknown length
39549 41479: contig of 1930 bp in length
41479 41579: gap of unknown length
41579 44370: contig of 2791 bp in length
44370 44470: gap of unknown length
44470 44771 44771: contig of 2113 bp in length
44771 46583: gap of unknown length
46583 46684 46684: contig of 4010 bp in length
46684 50793: gap of unknown length
50793 50794 50794: gap of unknown length
50794 53466: contig of 2673 bp in length
53466 53566: gap of unknown length
53566 56341: contig of 2775 bp in length
56341 56441: gap of unknown length
56441 58665: contig of 2224 bp in length
58665 58765: gap of unknown length
58765 61150: contig of 2385 bp in length
61150 61250: gap of unknown length
61250 63600: contig of 2350 bp in length
63600 63700: gap of unknown length
63700 66100: contig of 2400 bp in length
66100 66200: gap of unknown length
66200 68427: contig of 2227 bp in length
68427 68527: gap of unknown length
68527 71553: contig of 3026 bp in length
71553 71653: gap of unknown length
71653 74315: contig of 2662 bp in length
74315 74415: gap of unknown length
74415 77732: contig of 3317 bp in length
77732 77832: gap of unknown length
77832 80822: contig of 2990 bp in length
80822 80922: gap of unknown length
80922 84039: contig of 3117 bp in length
84039 84139: gap of unknown length
84139 87817: contig of 3678 bp in length
87817 87917: gap of unknown length
87917 91189: contig of 3272 bp in length
91189 91289: gap of unknown length

```

* 91290 96104: contig of 4815 bp in length
* 96105 96204: gap of unknown length
* 96205 101241: contig of 5037 bp in length
* 101242 101341: gap of unknown length
* 101342 106604: contig of 5263 bp in length
* 106605 106704: gap of unknown length
* 106705 111595: contig of 4891 bp in length
* 111596 111595: gap of unknown length
* 111597 115757: contig of 4062 bp in length
* 115758 115857: gap of unknown length
* 115858 121880: contig of 6023 bp in length
* 121881 121981: gap of unknown length
* 121982 126072: contig of 4091 bp in length
* 126072 126171: gap of unknown length
* 133330 133429: contig of 7158 bp in length
* 133430 138831: gap of unknown length
* 138832 138931: gap of 5402 bp in length
* 146739 146738: gap of unknown length
* 146839 146838: contig of 7807 bp in length
* 155333 155332: gap of unknown length
* 155433 155432: contig of 8494 bp in length
* 164997 164997: gap of unknown length
* 165098 165097: gap of 9565 bp in length
* 175000 175099: gap of unknown length
* 175100 187487: contig of 9902 bp in length
* 187487 187487: gap of unknown length

```

```

Query Match 70.0% Score 19.6; DB 2; Length 187487;
Best Local Similarity 84.6%; Pred. No. 99;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

3 TTCTGCGAGATGCGTCATGCGTGT 28
Db 51506 TTCTGCGAGATGCGTCATGCGTGT 51531

```

```

RESULT 4
AC122410 191844 bp DNA linear HTG 05-JUL-2002
LOCUS Mus musculus chromosome UNK clone RP24-142A1, WORKING DRAFT
DEFINITION AC122410
AC122410.2 GI:21699728
ACCESSION HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 191844)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 191844)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 191844)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (05-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Jul 5, 2002 this sequence version replaced gi:21105870.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M.BR0142A01

```

```

----- Summary Statistics -----
Sequencing vector: MJ3: 0%
Sequencing vector: plasmid: 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 188863 bases at least Q40
Consensus quality: 189311 bases at least Q20
Consensus quality: 189743 bases at least Q20
Insert size: 193000; agarose-1p
Insert size: 192017; sum-of-contigs
Quality coverage: 0.00 in Q20 bases; agarose-1p
Quality coverage: 11.99 in Q20 bases; sum-of-contigs

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1185: contig of 1185 bp in length
* 1186 1283: gap of unknown length
* 1286 2466: contig of 1181 bp in length
* 2467 2566: gap of unknown length
* 2567 8024: contig of 5458 bp in length
* 8025 8124: gap of unknown length
* 8125 18649: contig of 10525 bp in length
* 18650 18749: gap of unknown length
* 18750 42353: contig of 23604 bp in length
* 42354 42453: gap of unknown length
* 42454 82772: contig of 40319 bp in length
* 82773 82872: gap of unknown length
* 82873 191197: contig of 108325 bp in length
* 191198 191297: gap of unknown length
* 191298 191844: contig of 547 bp in length.

```

```

FEATURES
source
1. 191844
location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-142A1"
1. 1185

```

```

misc_feature /note="assembly_name:Contig22"
1286..2466
misc_feature /note="assembly_name:Contig23"
2367..8024
misc_feature /note="assembly_name:Contig24"
8125..18649
misc_feature /note="assembly_name:Contig25"
18750..42353
misc_feature /note="assembly_name:Contig26"
42454..82772
misc_feature /note="assembly_name:Contig27"
82873..191197
misc_feature /note="assembly_name:Contig28"
clone_end:sp6
vector_side:left"
191298..191844
/note="assembly_name:Contig8"
BASE COUNT 57770 a 38253 c 38232 g 56888 t 701 others
ORIGIN

```

```

Query Match 70.0% Score 19.6; DB 2; Length 191844;
Best Local Similarity 84.6%; Pred. No. 99;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

2 GTTCTGCGAGATGCGTCATGCGTGT 27
Db 153158 GTTCTGCGAGATGCGTCATGCGTGT 153183

```

```

RESULT 5
CET28A8      34388 bp   DNA       linear   INV 24-JAN-2002
LOCUS        Caenorhabditis elegans cosmid T28A8, complete sequence.
DEFINITION   292813
ACCESSION    292813.1 GI:3217752
VERSION      HTG.
KEYWORDS     Caenorhabditis elegans.
SOURCE       Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdittida;
ORGANISM     Rhabdittidae; Rhabdittidae; Pelodierinae; Caenorhabditis.

REFERENCE
AUTHORS      1
TITLE        none.
JOURNAL      Genome sequence of the nematode C. elegans: a platform for
MEDLINE      Investigating biology. The C. elegans Sequencing Consortium
PUBMED       Science 282 (5396), 2012-2018 (1998)
REMARK       2
REFERENCE    The C.elegans Sequencing Consortium.
AUTHORS      2 (bases 1 to 34388)
TITLE        Lloyd, C.R.
JOURNAL      Direct Submission
COMMENT      Submitted (09-MAR-1997) Nematode Sequencing Project, Sanger
              Institute, Hinxton, Cambridge CB10 1SA, England and Department of
              Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
              jesus@anger.ac.uk or rwenematode.wustl.edu
              On Jun 13, 1998 this sequence version replaced gi:1877160.
              Coding sequences below are predicted from computer analysis, using
              predictions from GeneFINDER (P. Green, U. Washington), and other
              available information.
              Current sequence finishing criteria for the C. elegans genome
              sequencing consortium are that all bases are either sequenced
              unambiguously on both strands, or on a single strand with both a
              dye primer and dye terminator reaction, from distinct subclones.
              Exceptions are indicated by an explicit note.
              IMPORTANT: This sequence is not the entire insert of clone T28A8.
              It may be shorter because we only sequence overlapping sections
              once, or longer because we arrange for a small overlap between
              neighbouring submissions.
              The true right end of clone T28A8 is at 104 in
              sequence AL032658.
              The true right end of clone Y43F4 is at 14666 in this sequence. The
              true right end of clone T7E9 is at 16930 in this sequence. The
              start of this sequence (1..104) overlaps with the end of sequence
              287059.
              The end of this sequence (34285..34388) overlaps with the start of
              sequence AL032658.
              For a graphical representation of this sequence and its analysis
              see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
              name=T28A8
              IMPORTANT: This sequence is NOT necessarily the entire insert of
              the specified clone. It may be shorter because we only sequence
              overlapping sections once, or longer because we arrange for a small
              overlap between neighbouring submissions.
              Location/Qualifiers
                1..34388
                  /organism="Caenorhabditis elegans"
                  /db_xref="taxon:6239"
                  /chromosome="T11"
                  /clone="T28A8"
                  /complement(join(2569..2659,2830..2990,3030..3099,
                  /gene="T28A8.1" 3235..3467,3861..3951,3998..4197,4279..4397,4889..5105))
                  /gene="T28A8.1" 3235..3467,3861..3951,3998..4197,4279..4397,4889..5105))
                  /complement(join(2569..2659,2830..2990,3030..3099,
                  /gene="T28A8.1" 3235..3467,3861..3951,3998..4197,4279..4397,4889..5105))
                  /note="contains similarity to Pfam domain: PF02034 (TRASK
                  k+ channel), Score=-80.4, E-value=3.4e-05, N=1"
                  /codon_start=1
                  /protein_id="CAB07284.1"
                  /db_xref="GI:3880336"
                  /db_xref="SPTREMBL:O9XU07"
                  /translation="MATWKTYYRIILAHVSLIVSVYVGGAFLFYOLEPNEVEVR

FEATURES
SOURCE
gene
CDS

```


Opportunistic pathogen
 JOURNAL Nature 406 (6799), 959-964 (2000)
 MEDLINE 20437337
 PUBMED 10984043
 REFERENCE 2 (bases 1 to 11884)
 AUTHORS Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Brinkman, F.S.L., Hutnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltzy, L., Tolentino, E., Westbrock-Madman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Lardig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Keller, J., Sater, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.
 TITLE Direct Submission
 JOURNAL Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
 FEATURES
 source Location/Qualifiers
 1. 11884
 /organism="Pseudomonas aeruginosa"
 /strain="PA01"
 /db_xref="taxon:287"
 102. 419
 /gene="PA3216"
 102. 419
 /gene="PA3216"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG06604.1"
 /db_xref="GI:9949337"
 /translation="MSSPDERFHSFAEFPYYLQESHNAVCRLHYGSLVLAIVF
 HAVASQOMLMLALPLVGYGPAWGVHFEKKNRATPKYPLMSFGMDVMKDAFTGR
 IR"
 636. 2027
 /gene="PA3217"
 636. 2027
 /gene="PA3217"
 /codon_start=1
 /transl_table=11
 /product="probable adenylate cyclase"
 /protein_id="AAG06605.1"
 /db_xref="GI:9949338"
 /translation="MKPTLPDRSPTSRSVSMREYSRVLAYIACGASIAAGTYQTF
 SYGLIMWVYALPLPHLAYHGORHODPKVTRALLAVDAVHCGMALIGSVYP
 SIMLVLTSFALYIGLRLIGMALIVASASALLAVIVAPILGNTPEVAASITF
 CGLYCITAFPGHOOGRLAQRQEIARBOEKARLARLNIAKITSFOYWEITEFSKKS
 VRLTQRKLLTFESDTRIGFELSELEAEALDNLNNYEMSKIAKIGTIDKTV
 GDCVAVFEGDPTSGAKKDAVAASVSMGIAMKRNKVLROQWRAQITRKPIRIGINT
 GYCTVGNFGADTRMDYTIIGREVMIAKRLSESASAGEILLISHEYTSILKIDVIMCRDKG
 QIAKGFSPVOIYQVVDSDRDLCAPSYVEHLPJFSMYLDNTNNIONDKERVIAL
 QQAERLRDRVYL"
 2569. 3042
 /gene="PA3218"
 2569. 3042
 /gene="PA3218"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG06606.1"
 /db_xref="GI:9949339"
 /translation="MEQFGIDALAPQVLEAPGAELVVAQARGAHRPAPMEGAOE
 GIAGGQRPRTAVQVFGKAMDRRGEOALAPAPRGPAERPLGGVIGIRREGID
 OPAEAPVPGGDRINAGTGAAEVGGRRHYLVAOQFOLALQGOAGDSDVDLR"
 complement(3069..3881)
 /gene="PA3219"
 complement(3069..3881)
 /gene="PA3219"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG06607.1"
 /db_xref="GI:9949340"
 /translation="MSTAQRTSNRKORVRLTISDVHLGTRDCAEHLAFLKRYHA

DRIVLGDIIIDGWLKRGIIYPOAHNTVIRLLTLKSRGTEVIYVTGNHDEFLRRSS
 LIIENIRLVDEAVSTADGRKLVIHGOQFQVITRYRHMLAFLDGSAVEFTLLNRL
 NHRRRNGCYGWSATLYKHVKRTAVNISDFEAAIAHECKRLDGVYCHTHAEI
 RRIEGVDILNCGDWESCTALIEHWDISIQILRAEQRALAAQAAAEVPA"
 4028. 4792
 /gene="PA3220"
 4028. 4792
 /gene="PA3220"
 /codon_start=1
 /transl_table=11
 /product="probable transcriptional regulator"
 /protein_id="AAG06608.1"
 /db_xref="GI:9949341"
 /translation="MGSILSRHYSHDOYVSHDAQVLGSGCLDPEVGRGSRVL
 RQFAVYPAQARHACSSPGRCLVLDPCGSLIGGEHAEKGRVLEPPOALSTP
 PAQOLYNWILASGPIINDPYIAHQAGALLASLASEROPPROQPLPLASLDATYDRHA
 AHPQVADLARLADLSVAREFSLTAEFGQPMQYAKSRRLHQAEHLLESIAVGEI
 AARYVASQSAFTYALVREFGVTPROLRRSRDXTG"
 4857. 5192
 /gene="csaA"
 /note="PA3221"
 4857. 5192
 /gene="csaA"
 /codon_start=1
 /transl_table=11
 /product="CsaA protein"
 /protein_id="AAG06609.1"
 /db_xref="GI:9949342"
 /translation="MDIIEQDFFERVELRVGTIISAAPNPKRPAVYLEVDLPGIL
 RTSQAQLTAHQAEQLVROVLCVNCNAPKQIAGVRESEVLTGVYDSEDKVYLAGFDK
 PLNGARLA"
 5202. 6101
 /gene="PA3222"
 5202. 6101
 /gene="PA3222"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG06610.1"
 /db_xref="GI:9949343"
 /translation="MNRSAALAHGALLGTLGVGFKLASAPGYIVGRALPAVL
 ALGLFARLAGAMRALGPAMQRLALGGLLGHVWSEFFIAVKNAGVAIVAGFASPE
 AFYVILDELIFERIRHIAEGLLVVLVGLVLPNDGLAGATLGLMALISGLLEFA
 LLSIANNSGQGVQVQALQNNVVALCLLPFVSELPALPDMLMIALIGVCGG
 LASHLTVASLAVVAPRAIYFALEPIYGIYFAMLLPHETGPRMFLGALITIAIYL
 SARIGANKRTAVDEKDEPERA"
 complement(6206..6847)
 /gene="acpD"
 /note="PA3223"
 complement(6206..6847)
 /gene="acpD"
 /codon_start=1
 /transl_table=11
 /product="acyl carrier protein phosphodiesterase"
 /protein_id="AAG06611.1"
 /db_xref="GI:9949344"
 /translation="MSRVIVTESSARQGSVRLTAFTSHMTIAPADQFOYRDLA
 REPLPHDELILGAWTTPCGHSAERARLERSRLTEELRMAVLYLAAPMTPALP
 SSLKSWFDVLRAGLTFRTYAOGEGLQGRARVLTARGGIVAGGGLDQEPYLRQV
 LGFVGIDVTFPIHAGMMGPEFEKGLARAREMRQALFDTSLCYPLPLR"
 complement(6994..7311)
 /gene="PA3224"
 complement(6994..7311)
 /gene="PA3224"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG06612.1"
 /db_xref="GI:9949345"
 /translation="MKRLCCPALPLPSAFAYIEVDPQNLGTEVSYTODIRDMK
 AILTLTNLGGQAAECTAVFRNGPELPVKRGVIGQGNNSFTAQGRDIIKRLIKLTCK
 IAK"
 7511. 8440

CDS

```

/gene="PA3225"
/511..8440
/gene="PA3225"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="AA06614.1"
/db_xref="GI:9949346"
/translation="MKAPVTIDQMTQAVVHGGPQAQAALHNSQSSISYTVARM
OEOLGVPLRIDGRKAVITERGVEYLRSQVLKQAGLELAHMEQWPEPVRLV
DAAYPTARIYVLSAPFQSGCVLLREVLSGVEALHOGAADIATSLITPGILG
ADLSVFEVAVAHDPPLHLRLRELTQHDDETOMQVYIRDSGLQPRDHGMGAERW
TWGSLATAATFVGNGEFAWLPRIITERELOSQLPPLPSGCVQNSFYLTNKEK
PLGPAQSLIMEMKSPFANVPLNAPFAPEPAAD"
8447..9274
/gene="PA3226"
/8447..9274
/gene="PA3226"
/codon_start=1
/transl_table=11
/product="probable hydrolase"
/protein_id="AA06614.1"
/db_xref="GI:9949347"
/translation="MPDFQROCTIVYDDHDPGTERAPVLLHGLSSARDWEYOLPA
LIGRYRLVPLDNGHSGSKPGGYSMAGFAEDCALDLRLCGPVHVLGSMGMIG
FOLACRPLRLSRITVNSTPEYTPKPREHELVAKRKLISLRLTIGRALGILF
PKAEQGLIKETKFAHAEKNDKRAVLSADAIIGWQVIERLAIACPTVVISADRYTP
VSLKQATCERILDLARLVEDSRHATPMDQPIFTMTLLGLADIKREQCSVVS"
9256..9819
/gene="p1A"
/1011..1121
/1011..1121
/gene="p1A"
/codon_start=1
/transl_table=11
/product="peptidyl-prolyl cis-trans isomerase A"
/protein_id="AA06615.1"

```

gene

```

/gene="p1A"
/1011..1121
/1011..1121
/gene="p1A"
/codon_start=1
/transl_table=11
/product="peptidyl-prolyl cis-trans isomerase A"
/protein_id="AA06615.1"

```

Query Match

Best local similarity 67.9% Score 19; DB 1; Length 11884;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

DB

1457 GTTCGTGGGAGTCCGTCATGCTGT 28
1457 GTTCGTGGGAGTCCGTCATGCTGT 1463

RESULT 8

AC126800/c 14687 bp DNA linear HTG 09-JUL-2002

LOCUS

DEFINITION Mus musculus chromosome UNK clone RP24-447D19, WORKING DRAFT

ACCESSION AC126800

VERSION AC126800.1 GI:21717270

KEYWORDS HTG; HTGS; PHASE1; HTGS; DRAFT.

SOURCE house mouse

ORGANISM Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (09-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>

Contact: submissions@wustl.edu
Project information
Center project name: W_BB0447D19

Summary Statistics

Sequencing vector: M13; 0%

Sequencing vector: plasmid; 100%

Chemistry: Dye-primer ET; 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 11885 bases at least Q40

Consensus quality: 12505 bases at least Q30

Consensus quality: 13093 bases at least Q20

Insert size: 175000; agarose-fp

Insert size: 13787; sum-of-contigs

Quality coverage: 0.12 in Q20 bases; agarose-fp

Quality coverage: 1.48 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1121: contig of 1121 bp in length

1122 1221: gap of unknown length

1222 2542: contig of 1321 bp in length

2543 2643: gap of unknown length

2643 3696: contig of 1053 bp in length

3697 3797: gap of unknown length

3797 5103: contig of 1307 bp in length

5104 5203: gap of unknown length

5204 6395: contig of 1192 bp in length

6396 6496: gap of unknown length

6496 8013: contig of 1518 bp in length

8014 8114: gap of unknown length

8114 10012: contig of 1899 bp in length

10013 10113: gap of unknown length

10113 11603: contig of 1491 bp in length

11604 11703: gap of unknown length

11704 13134: contig of 1431 bp in length

13135 13235: gap of unknown length

13235 14687: contig of 1453 bp in length.

FEATURES

Source

1..14687

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="UNK"

/clone="RP24-447D19"

misc_feature

1..1121

/note="assembly_name:Contig15"

1222..2542

/note="assembly_name:Contig31"

2643..3696

/note="assembly_name:Contig33"

3797..5103

/note="assembly_name:Contig35"

5204..6395

/note="assembly_name:Contig38"

6496..8013

/note="assembly_name:Contig40"

8114..10012

/note="assembly_name:Contig41"

10113..11603

/note="assembly_name:Contig42"

11704..13134

/note="assembly_name:Contig43"

13235..14687

/note="assembly_name:Contig44"

misc_feature

misc_feature

misc_feature

BASE COUNT 3528 a 3262 c 3191 g 3797 t 909 others

ORIGIN


```

repeat_region      /rpt_family="AT_rich"
                    14527..14866
repeat_region      /rpt_family="ERV1"
                    14867..14965
repeat_region      /rpt_family="ERV1"
                    16820..17126
repeat_region      /rpt_family="Alu"
                    18043..18394
repeat_region      /rpt_family="ERV1"
                    18412..18744
repeat_region      /rpt_family="ERV1"
                    18745..18789
repeat_region      /rpt_family="ERV1"
                    18836..18900
repeat_region      /rpt_family="TTCGCCn"
                    18901..18966
repeat_region      /rpt_family="ERV1"
                    18968..19124
repeat_region      /rpt_family="MERL_type"
                    20267..20317
repeat_region      /rpt_family="AT_rich"
                    20490..20538
repeat_region      /rpt_family="AT_rich"
                    21037..21057
repeat_region      /rpt_family="AT_rich"
                    21576..22017
repeat_region      /rpt_family="L2"
                    22184..22548
repeat_region      /rpt_family="MALR"
                    22561..22868
repeat_region      /rpt_family="MALR"
                    24085..24256
repeat_region      /rpt_family="L2"
                    24283..24402
repeat_region      /rpt_family="L1"
                    24404..24491
repeat_region      /rpt_family="MERS3"
                    24503..25061
repeat_region      /rpt_family="L2"
                    25067..25137
repeat_region      /rpt_family="L2"
                    25166..26022
repeat_region      /rpt_family="L2"
                    26239..26260
repeat_region      /rpt_family="AT_rich"
                    26472..26500
repeat_region      /rpt_family="AT_rich"
                    26870..26898
repeat_region      /rpt_family="AT_rich"
                    27957..28030
repeat_region      /rpt_family="AT_rich"
                    28166..28189
repeat_region      /rpt_family="AT_rich"
                    28670..28714
repeat_region      /rpt_family="AT_rich"
                    28745..29121
repeat_region      /rpt_family="L2"
                    29265..29563
repeat_region      /rpt_family="ERV1"
                    29643..29778
repeat_region      /rpt_family="MIR"
                    29813..29856
repeat_region      /rpt_family="AT_rich"
                    30211..30507
repeat_region      /rpt_family="Alu"
                    30839..31786

```

Query Match 67.9%; Score 19; DB 9; Length:80015;
 Best Local Similarity 81.5%; Pred. No. 1.9e+02;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

2 GTTCGTGGATAGTCGCTGATGTT 28
 ||||| ||||| |||||

Db 54225 GTTCGTGGATAGTCGCTGATGTT 54199

RESULT 10
 AC004414/c 115045 bp DNA linear PRI 21-DEC-1999
 LOCUS Homo sapiens PAC clone RP4-724E13 from 7p11.2-p12, complete
 DEFINITION
 ACCESSION AC004414
 VERSION AC004414
 KEYWORDS HGC
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS Smith A and Le T
 TITLE The sequence of Homo sapiens PAC clone RP4-724E13
 JOURNAL Unpublished
 2 (bases 1 to 115045)
 WATERSTON, R.
 DIRECT SUBMISSION
 Submitted (20-MAR-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 3 (bases 1 to 115045)
 WATERSTON, R.
 DIRECT SUBMISSION
 Submitted (21-DEC-1999) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE
 AUTHORS Smith A and Le T
 TITLE The sequence of Homo sapiens PAC clone RP4-724E13
 JOURNAL Unpublished
 2 (bases 1 to 115045)
 WATERSTON, R.
 DIRECT SUBMISSION
 Submitted (20-MAR-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 3 (bases 1 to 115045)
 WATERSTON, R.
 DIRECT SUBMISSION
 Submitted (21-DEC-1999) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

Center: Washington University Genome Sequencing Center
 Web site: http://genome.wustl.edu/gsc
 Contact: sapient@wustl.wustl.edu
 Center project name: H_DJ0724E13

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate
 chemistry; an attempt was made to resolve all sequencing problems,
 such as compressions and repeats; all regions were covered by
 sequence from more than one subclone; and the assembly was
 confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and
 sequencing collaboration between the NHGRI Chromosome 7 Mapping
 Project (Eric D. Green, Director), John D. McPherson in the
 Department of Genetics (Washington University), and the Washington
 University Genome Sequencing Center. For additional information
 about the map position of this sequence, see
 http://www.nhgri.nih.gov/DIR/GRN/CH7 send
 mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
 This clone was derived from human PAC library RPI-4, prepared by
 Pliet de Jong and coworkers at Roswell Park Cancer Institute,
 using the method described by Ioannou et al., Nature Genetics
 6:84-9 (1994). The library is from one male donor. For further
 details, see http://bacpac.med.buffalo.edu/
 The clone is available from Genome Systems, Inc.
 (http://www.genomesystems.com).

NEIGHBORING SEQUENCE INFORMATION:
 This clone sequenced to the left is RP4-537J23. Actual start of
 this clone is at base position of 1 of RP4-724E13; actual end is at
 115045 of RP4-724E13.

FEATURES This clone contains STS SWS3034 (NTD:G1233489).
 source Location/Qualifiers

1. 115045
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7p11.2-p12"
 /clone="RP4-724E13"
 /clone_id="RPC1-4"
 115. 211
 /rpt_family="MIR"
 785. 1150
 /rpt_family="MALR"
 1151. 1286
 /rpt_family="MER1_type"
 1518. 1828
 /rpt_family="L2"
 1845. 2157
 /rpt_family="Alu"
 3529. 3585
 /rpt_family="MIR"
 3862. 3950
 /rpt_family="MIR"
 4004. 4351
 /rpt_family="MIR"
 4854. 5198
 /rpt_family="MER51A"
 5197. 5608
 /rpt_family="L1"
 6177. 6227
 /rpt_family="L1"
 6229. 6384
 /rpt_family="MER4-group"
 6397. 6633
 /rpt_family="MER4-group"
 6646. 6738
 /rpt_family="MER4-group"
 7798. 7919
 /rpt_family="MER4-group"
 7937. 8097
 /rpt_family="MER4-group"
 8089. 8568
 /rpt_family="MER4-group"
 8564. 8935
 /rpt_family="MER4-group"
 8937. 9028
 /rpt_family="MER4-group"
 9025. 9461
 /rpt_family="MER4-group"
 9471. 9512
 /rpt_family="MER4-group"
 9600. 9677
 /rpt_family="L2"
 10659. 10961
 /rpt_family="Alu"
 11884. 12003
 /rpt_family="L1"
 12172. 12299
 /rpt_family="L1"
 12765. 12935
 /rpt_family="Alu"
 14936. 15312
 /rpt_family="L1"
 15680. 16117
 /rpt_family="MALR"
 16252. 16430
 /rpt_family="MALR"
 16432. 16727
 /rpt_family="L1"
 16744. 16809
 /rpt_family="Alu"
 16812. 17261
 /rpt_family="L1"
 repeat_region

repeat_region /rpt_family="MER4-group"
 17262. 17711
 /rpt_family="L1"
 repeat_region 17742. 18017
 /rpt_family="Alu"
 repeat_region 18646. 18855
 /rpt_family="MER1_type"
 repeat_region 19218. 19709
 /rpt_family="L1"
 repeat_region 22186. 22273
 /rpt_family="MIR"
 repeat_region 22826. 23017
 /rpt_family="Retroviral"
 repeat_region 23020. 23168
 /rpt_family="Alu"
 repeat_region 23412. 23456
 /rpt_family="Retroviral"
 repeat_region 23491. 23642
 /rpt_family="MIR"
 repeat_region 23820. 24014
 /rpt_family="MER1_type"
 repeat_region 24355. 24609
 /rpt_family="MIR"
 repeat_region 24903. 25536
 /rpt_family="MER4-group"
 repeat_region 26083. 26644
 /rpt_family="L1"
 repeat_region 26846. 26947
 /rpt_family="Alu"
 repeat_region 26979. 27362
 /rpt_family="L1"
 repeat_region 27449. 28151
 /rpt_family="L1"
 repeat_region 28223. 28485
 /rpt_family="MER2_type"
 repeat_region 28487. 29298
 /rpt_family="MER21_g"
 repeat_region 29312. 29978
 /rpt_family="L1"
 repeat_region 29990. 30366
 /rpt_family="L1"
 repeat_region 30351. 32031
 /rpt_family="L1"
 repeat_region 32030. 32521
 /rpt_family="L1"
 repeat_region 32536. 32929
 /rpt_family="Retroviral"
 repeat_region 32937. 33598
 /rpt_family="L1"
 repeat_region 33639. 33939
 /rpt_family="Alu"
 repeat_region 34345. 34628
 /rpt_family="L1"
 repeat_region 35643. 35945
 /rpt_family="Alu"
 repeat_region 35986. 36474
 /rpt_family="L1"
 repeat_region 36497. 36719
 /rpt_family="L1"

Query Match 67.9% Score 19; DB 9; Length 115045;
 Best Local Similarity 81.5% Pred. No. 2e+02;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 GTTCGTGGATGATCGTATGCTT 28
 DB 41511 GTTCGTGGATGATCGTATGCTT 41485

RESULT 11
 AC120560 123770 bp DNA linear HTG 18-JUL-2002
 LOCUS AC120560/c
 DEFINITION Rattus norvegicus clone CH230-330L5, *** SEQUENCING IN PROGRESS

ACCESSION
AC120560
VERSION
AC120560.2 GI:21745753
KEYWORDS
HTG: HTGS PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus

REFERENCE
AUTHORS

***, 29 unordered pieces.
Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C., Alstrocks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,T., Benton,J., Blumhage,K., Blankenburg,K., Bonin,D., Bouck,J., Bowler,S., Brileva,M., Brown,E., Brown,M., Bryant,D., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gadioli,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,S., Huliyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lileu,C., Liu,J., Liu,W., Lohr,C., Lozad,R.J., Lu,X., Lucier,A., Lucier,K., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massay,E., Mahoney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,M., Morris,S., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunnu,G., Oragunye,N., Oyedero,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,J., Plickens,R., Primus,E., Pu,L.L., Qulies,M., Ren,Y., Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H., Taneey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlaczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

TITLE

Unpublished

JOURNAL

21 (bases 1 to 123770)

REFERENCE

Submitted (09-MAY-2002)

AUTHORS

Worley,K.C.

JOURNAL

Submitted (09-MAY-2002)

REFERENCE

Submitted (09-MAY-2002)

AUTHORS

Worley,K.C.

JOURNAL

Submitted (09-MAY-2002)

REFERENCE

Submitted (09-MAY-2002)

AUTHORS

Worley,K.C.

JOURNAL

Submitted (09-MAY-2002)

REFERENCE

Submitted (09-MAY-2002)

AUTHORS

Worley,K.C.

JOURNAL

Submitted (09-MAY-2002)

REFERENCE

Submitted (09-MAY-2002)

AUTHORS

Worley,K.C.

JOURNAL

Submitted (09-MAY-2002)

REFERENCE

Submitted (09-MAY-2002)

AUTHORS

Worley,K.C.

JOURNAL

Submitted (09-MAY-2002)

REFERENCE

Submitted (09-MAY-2002)

AUTHORS

Worley,K.C.

JOURNAL

Submitted (09-MAY-2002)

REFERENCE

Submitted (09-MAY-2002)

AUTHORS

Worley,K.C.

JOURNAL

Submitted (09-MAY-2002)

REFERENCE

Submitted (09-MAY-2002)

AUTHORS

Worley,K.C.

JOURNAL

Submitted (09-MAY-2002)

Summary Statistics
Sequencing vector: plasmid;
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 96476 bases at least Q40
Consensus quality: 99583 bases at least Q30
Consensus quality: 101766 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 29 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1	1038:	contig of 1038 bp in length
1039	1138:	gap of unknown length
1139	2213:	contig of 1075 bp in length
2214	2313:	gap of unknown length
2314	3496:	contig of 1183 bp in length
3497	3596:	gap of unknown length
3597	4962:	contig of 1366 bp in length
4963	6379:	contig of 1317 bp in length
6380	6479:	gap of unknown length
6480	8102:	contig of 1623 bp in length
8103	8202:	gap of unknown length
8203	9510:	contig of 1308 bp in length
9511	9610:	gap of unknown length
9611	10987:	contig of 1377 bp in length
10988	11087:	gap of unknown length
11088	12208:	contig of 1121 bp in length
12209	12308:	gap of unknown length
12309	13660:	contig of 1332 bp in length
13661	13760:	gap of unknown length
13761	15626:	contig of 1866 bp in length
15627	15726:	gap of unknown length
15727	17972:	contig of 2246 bp in length
17973	18072:	gap of unknown length
18073	20240:	contig of 2168 bp in length
20241	20340:	gap of unknown length
20341	23369:	contig of 2929 bp in length
23370	25441:	contig of 2072 bp in length
23370	25441:	gap of unknown length
25442	25541:	gap of unknown length
25542	29079:	contig of 3538 bp in length
29080	29179:	gap of unknown length
29180	33526:	contig of 4347 bp in length
33527	33626:	gap of unknown length
33627	37047:	contig of 3421 bp in length
37048	37147:	gap of unknown length
37148	40494:	contig of 3347 bp in length
40495	40594:	gap of unknown length
40595	44084:	contig of 3490 bp in length
44085	44184:	gap of unknown length
44185	48306:	contig of 4122 bp in length
48307	48406:	gap of unknown length
48407	53058:	contig of 4652 bp in length
53059	53158:	gap of unknown length
53159	60470:	contig of 7312 bp in length
60471	60570:	gap of unknown length
60571	67470:	contig of 6900 bp in length
67471	67570:	gap of unknown length
67571	72023:	contig of 4453 bp in length
72024	72123:	gap of unknown length
72124	82882:	contig of 10759 bp in length
82883	82982:	gap of unknown length
82983	89105:	contig of 6123 bp in length
89106	89205:	gap of unknown length
89206	98084:	contig of 8879 bp in length

```

* 98085 98184: gap of unknown length
* 98185 123770: contig of 25586 bp in length.
FEATURES
SOURCE
1. 123770
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-33015"
BASE COUNT 33762 a 24543 c 24266 g 33085 t 8114 others
ORIGIN
Query Match 67.9%; Score 19; DB 2; Length 123770;
Best Local Similarity 81.5%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 2 GTGCGGATAGTCGTCATGCGTGT 28
DB 48045 GTTCCTGGGTAGTCATCAACGTGT 48019

RESULT 12
AC094593/C
LOCUS
DEFINITION AC094593 153987 bp DNA linear HTG 10-JUL-2002
Rattus norvegicus clone CH230-4D11, *** SEQUENCING IN PROGRESS ***
ACCESSION AC094593
VERSION AC094593.3 GI:21716291
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 153987)
Munzy,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,
Alsbrooks,S.E., Amaralunga,H.C., Ate,J.R., Ayale,M., Banks,T.,
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonin,D.,
Bouck,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,U., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsl,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kritovyc,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
Lorado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Mettler,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,A., Morris,S.,
Moser,M., Nickerson,E., Nwokenwo,S., Oguh,M., Okunolu,G.,
Nguyen,N., Nguyen,N., Nguyen,N., Nguyen,N., Nguyen,N.,
Ongunye,N., Oviado,R., Pace,A., Payton,L., Peery,J., Perez,L.,
Peters,L., Pickens,R., Pace,A., Payton,L., Peery,J., Perez,L.,
Peters,L., Pickens,R., Pace,A., Payton,L., Peery,J., Perez,L.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Syarik,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,K.,
Williams,G., Williamson,A., Wlarczyk,R., Wooden,S., Wolley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Direct Submission
Unpublished
2 (bases 1 to 153987)
Worley,K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 153987)
Worley,K.C.
Direct Submission
Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 9, 2002 this sequence version replaced gi:17941355.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GAW
Center clone name: CH230-4D11
Sequencing vector: plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap version 0.990329
Consensus quality: 95495 bases at least Q40
Consensus quality: 102767 bases at least Q30
Consensus quality: 108492 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 75 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1087 1086: contig of 1086 bp in length
1187 1186: gap of unknown length
2490 2490: contig of 1304 bp in length
2491 2590: gap of unknown length
3752 3752: contig of 1152 bp in length
3753 3852: gap of unknown length
3853 4943: contig of 1091 bp in length
4944 5043: gap of unknown length
5044 6513: contig of 1470 bp in length
6514 6613: gap of unknown length
6614 7864: contig of 1251 bp in length
7865 7964: gap of unknown length
7965 9271: contig of 1307 bp in length
9272 9371: gap of unknown length
9372 10391: contig of 1020 bp in length
10392 10491: gap of unknown length
10492 11566: contig of 1075 bp in length
11567 11666: gap of unknown length
11667 13142: contig of 1476 bp in length
13143 13242: gap of unknown length
13243 14800: contig of 1558 bp in length
14801 14900: gap of unknown length
14901 16133: contig of 1233 bp in length
16134 16233: gap of unknown length
16234 17334: contig of 1101 bp in length
17335 17434: gap of unknown length
17435 18437: contig of 1003 bp in length
18438 18537: gap of unknown length
18538 19964: contig of 1427 bp in length
19965 20064: gap of unknown length
20065 21617: contig of 1553 bp in length
21617 21618: gap of unknown length

```

	Query Match	Best Local Similarity	Matches	Conservative	Score 19;	DB 2;	Length 153987;	Pred. No. 2e+02;	Mismatches	Indels	Gaps
	*	88350	88359	gap of unknown length							
		88320	90225	confli of 1876 bp in length							
Oy	2	GTTCGATGGGATATGCGCGTATAGGTGTT	28								
Db	77771	GTTCGATGGGATATCTTCCTCGGCGTT	77745								

RESULT 13	AC117647/C	LOCUS	DEFINITION	AC117647	179572 bp	DNA	linear	HTG_06-AUG-2002
			Mus musculus clone RP23-244M7, WORKING DRAFT SEQUENCE, 22 ordered pieces.	AC117647				
ACCESSION	AC117647			AC117647				
VERSION	AC117647.2			GI:22123324				
KEYWORDS	HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.							
SOURCE	house mouse.							
ORGANISM	Mus musculus							

REFERENCE AUTHORS	TITLE	JOURNAL	REFERENCE AUTHORS
1 (bases 1 to 179572)			
Birren, B., Nusbaum, C. and Lander, E.	Mus musculus, clone RP23-244M7	Unpublished	
2 (bases 1 to 179572)			
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, A.,...			

TITLE Direct Submission of 2D and 3D Data
JOURNAL Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome Research, 330 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (phases 1 to 179572)
AUTHORS Birren, B., Nusbaum, C., Lander, E., Ali, A., Nicio, V., ...

Camarta, J., Chang, J., Broom, T., Boguslavsky, L., Boukragter, B., Cook, A., Cooke, P., Chazaro, B., Choepel'y, L., Collimore, A., Fero, S., Ferreira, P., DeArillano, K., Dewar, K., Diaz, J.S., Dodge, S., Gardyaa, S., Gord, S., Fitzgerald, M., Gage, D., Galagan, J., Horton, L., Hulme, W., Graham, L., Grand-Pierre, N., Hagos, B., Karatas, A., Kells, C., Landers, T., Levine, R., Jones, C., Kamat, A., Llu'g, Maclean, C., Macdonald, P., Major, J., Matthews, C., McCathy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhking, P., Pierre, N., Raymond, C., Retter, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuppach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Toham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Ziembeck, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 6, 2002 this sequence version replaced g1:20128406.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W1BR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L23595

Center clone name: 244.M.7

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 172421 bases at least Q40
Consensus quality: 175743 bases at least Q30
Consensus quality: 176806 bases at least Q20
Insert size: 182000; agarose-fp
Quality coverage: 5.0 in Q20 bases; agarose-fp
Quality coverage: 5.1 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```

1      831: contig of 831 bp in length
*      832 931: gap of 100 bp
*      932 2184: contig of 1253 bp in length
*      2185 2284: gap of 100 bp
*      2285 3205: contig of 921 bp in length
*      3206 3305: gap of 100 bp
*      3306 4649: contig of 1344 bp in length
*      4650 4749: gap of 100 bp
*      4750 7578: contig of 2829 bp in length
*      7579 7678: gap of 100 bp
*      7679 10016: contig of 2338 bp in length
*      10017 10116: gap of 100 bp
*      10117 11576: contig of 1460 bp in length
*      11577 11676: gap of 100 bp
*      11677 14279: contig of 2603 bp in length
*      14280 14379: gap of 100 bp
*      14380 52253: contig of 37874 bp in length
*      52254 52353: gap of 100 bp
*      52354 56294: contig of 3941 bp in length
*      56295 56394: gap of 100 bp
*      56395 62939: contig of 6545 bp in length
*      62940 63039: gap of 100 bp
*      63040 67332: contig of 4293 bp in length
*      67333 67432: gap of 100 bp
*      67433 74902: contig of 7470 bp in length
*      74903 75002: gap of 100 bp
*      75003 85025: contig of 10023 bp in length
*      85026 85125: gap of 100 bp
*      85126 91380: contig of 6555 bp in length
*      91381 91480: gap of 100 bp
*      91481 98876: contig of 7396 bp in length
*      98877 98976: gap of 100 bp
*      98977 106865: contig of 7889 bp in length
*      106866 106965: gap of 100 bp
*      106966 119527: contig of 12562 bp in length
*      119528 119627: gap of 100 bp
*      119628 135356: contig of 15729 bp in length

```

```

*      135357 135456: gap of 100 bp
*      135457 152890: contig of 17434 bp in length
*      152891 152990: gap of 100 bp
*      152991 176149: contig of 22159 bp in length
*      176150 176249: gap of 100 bp
*      176250 179572: contig of 3323 bp in length.
      location/Qualifiers
1. 179572

```

FEATURES

source

misc_feature

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-244M7"
/clone_11b="RPC1-23 Female Mouse BAC"
1. 831

```

misc_feature

```

/clone_end:SP6
vector_side:left"

```

misc_feature

```

/organism="assembly_fragment"
2285. 3205

```

misc_feature

```

/organism="assembly_fragment"
3306. 4649

```

misc_feature

```

/organism="assembly_fragment"
4750. 7578

```

misc_feature

```

/organism="assembly_fragment"
7679. 10016

```

misc_feature

```

/organism="assembly_fragment"
10117. 11576

```

misc_feature

```

/organism="assembly_fragment"
11677. 14279

```

misc_feature

```

/organism="assembly_fragment"
14380. 52253

```

misc_feature

```

/organism="assembly_fragment"
52354. 56294

```

misc_feature

```

/organism="assembly_fragment"
56395. 62939

```

misc_feature

```

/organism="assembly_fragment"
63040. 67332

```

misc_feature

```

/organism="assembly_fragment"
67433. 74902

```

misc_feature

```

/organism="assembly_fragment"
75003. 85025

```

misc_feature

```

/organism="assembly_fragment"
85126. 91380

```

misc_feature

```

/organism="assembly_fragment"
91481. 98876

```

misc_feature

```

/organism="assembly_fragment"
98977. 106865

```

misc_feature

```

/organism="assembly_fragment"
106966. 119527

```

misc_feature

```

/organism="assembly_fragment"
119628. 135356

```

misc_feature

```

/organism="assembly_fragment"
135457. 152890

```

misc_feature

```

/organism="assembly_fragment"
152991. 176149

```

misc_feature

```

/organism="assembly_fragment"
176250. 179572

```

misc_feature

```

/organism="assembly_fragment"
vector_end:77
vector_side:right"

```

BASE COUNT 55686 a 32751 c 33323 g 55711 t 2101 others

ORIGIN

Query Match

Best Local Similarity 67.9%; Score 19; DB 2; Length 179572;

Matches 22; Conservativity 0; Mismatches 5; Indels 0; Gaps 0;

2 GTTGTGGATGACGCGATGCGT 28

DB 158787 GTTCTTGGAAAGTACATGCTGT 158761

RESULT 14

AC009921
LOCUS
DEFINITION
AC009921 Homo sapiens clone RP11-115018, WORKING DRAFT SEQUENCE, 16
AC009921
AC009921.4 GI:7329301
HTG: HTGS_PHASE1, HTGS_DRAFT
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
Unpublished
2 (bases 1 to 189086)
Britten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baker, J., Baldwin, J., Barna, N., Beckert, R., Benn, J., Brown, A.,
Caste, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeRellano, K., Depayre, E., Devon, K., Dewar, K.,
Donelan, L., Doyle, M., Ferrel, P., Fitzhugh, W., Forrest, C.,
Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,
Karatas, A., Lehoczy, J., Lieu, C., Locke, K., MacDonald, P.,
Marquis, N., McGowan, P., McGurk, A., McKernan, K., McLaughlin, J.,
Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
Stevale, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
Wiesler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
Submitted (08-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6479138.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
JOURNAL
COMMENT
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: 11914
Center clone name: 115.O.18

Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 179818 bases at least Q40
Consensus quality: 183946 bases at least Q30
Consensus quality: 185564 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 187586; sum-of-ctnigs
Quality coverage: 5.2 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-ctnigs

NOTE: This is a 'working draft' sequence. It currently
consists of 16 ctnigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the ctnigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1064::contig of 1064 bp in length
1065 1164::gap of 100 bp
1165 2211::contig of 1047 bp in length
2212 2311::gap of 100 bp
2312 3615::contig of 1304 bp in length
3616 3715::gap of 100 bp
3716 5319::contig of 1604 bp in length

FEATURES
Source
* 5320 5419:: gap of 100 bp
* 5420 7668:: contig of 2249 bp in length
* 7669 7768:: gap of 100 bp
* 7769 12506:: contig of 4738 bp in length
* 12507 12606:: gap of 100 bp
* 12607 17063:: contig of 4457 bp in length
* 17064 17163:: gap of 100 bp
* 17164 24729:: contig of 7566 bp in length
* 24730 24829:: gap of 100 bp
* 24830 35834:: contig of 11005 bp in length
* 35835 35934:: gap of 100 bp
* 35935 46466:: contig of 10532 bp in length
* 46467 46566:: gap of 100 bp
* 46567 61624:: contig of 15058 bp in length
* 61625 61724:: gap of 100 bp
* 61725 77509:: contig of 15785 bp in length
* 77510 77609:: gap of 100 bp
* 77610 94997:: contig of 17388 bp in length
* 94998 95097:: gap of 100 bp
* 95098 120488:: contig of 23391 bp in length
* 120489 120588:: gap of 100 bp
* 120589 147913:: contig of 27325 bp in length
* 147914 148013:: gap of 100 bp
* 148014 189086:: contig of 41073 bp in length.
Location/Qualifiers
1. 189086
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-115018"
/clone_1bp="RPCT-11 Human Male BAC"

misc_feature
1. 1064
/note="assembly-fragment"
misc_feature
1165. 2211
/note="assembly-fragment"
misc_feature
2312. 3615
/note="assembly-fragment"
misc_feature
3716. 5319
/note="assembly-fragment"
misc_feature
5420. 7668
/note="assembly-fragment"
misc_feature
7769. 12506
/note="assembly-fragment"
misc_feature
12607. 17063
/note="assembly-fragment"
misc_feature
17164. 24729
/note="assembly-fragment"
misc_feature
24830. 35834
/note="assembly-fragment"
misc_feature
35935. 46466
/note="assembly-fragment"
misc_feature
46567. 61624
/note="assembly-fragment"
misc_feature
61725. 77509
/note="assembly-fragment"
misc_feature
77610. 94997
/note="assembly-fragment"
misc_feature
95098. 120488
/note="assembly-fragment"
misc_feature
120589. 147913
/note="assembly-fragment"
misc_feature
148014. 189086
/note="assembly-fragment"
BASE COUNT 61951 a 33004 c 33086 g 59542 t 1503 others
ORIGIN

Query Match 67.9%; Score 19; DB 2; Length 189086;
Best Local Similarity 81.5%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 GTTCGTGGATAGTCGTCATGCTGT 28
 Db 88210 GTTCCTGGTATGTCATGCTGT 88236
 RESULT 15
 AL354896/C
 LOCUS
 DEFINITION Human DNA sequence from clone RP11-512M17 on chromosome 13,
 complete sequence.
 ACCESSION AL354896 209743 bp DNA linear PRI 29-NOV-2000
 VERSION AL354896
 KEYWORDS AL354896.16 GI:11322813
 SOURCE HTG.
 ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 209743)
 AUTHORS Lovell, J.
 TITLE Direct Submission
 JOURNAL Submitted (29-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 Requests: clone@sanger.ac.uk
 On Nov 23, 2000 this sequence version replaced gi:113221907.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated repeat sequence elements. Where the sequence is
 ambiguous, there is an annotation using the 'unSURE' feature key.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em., EMBL; SW, SWISSPROT; Tr., TrEMBL; Wp., WormBase; Information
 on the WormBase database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormbase
 This sequence was generated from part of bacterial clone contigs of human
 chromosome 13, constructed by the Sanger Centre Chromosome 13
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr13
 RP11-512M17 is from the library RPCR-11.2 constructed at the
 Roswell Park Cancer Institute by the group of Pieter de Jong. For
 further details see http://bacpac.med.buffalo.edu/
 VECTOR: pBAC3.6
 This sequence is the entire insert of clone RP11-512M17 The true
 left end of clone RP11-545P6 is at 121889 in this sequence. The
 true right end of clone RP11-315A9 is at 159267 in this sequence.
 FEATURES
 SOURCE
 Location/Qualifiers
 1..209743.
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /clone="RP11-512M17"
 /clone_lib="RPCR-11.2"
 1..265
 /note="MSTR repeat: matches 160. 426 of consensus"
 repeat_region 270..1579
 /note="L1P12 repeat: matches -55. 1238 of consensus"
 repeat_region 1572..2157
 /note="L1P12 repeat: matches -1416. -839 of consensus"
 repeat_region 2153..4016
 /note="L1 repeat: matches 2124. 4006 of consensus"
 misc_feature 2830..2867
 /note="Sequence from overlapping clone
 BA315A9(AL162498). Assembly confirmed by restriction
 digest."
 repeat_region 3946..4064
 /note="L1M2 repeat: matches 2455. 2571 of consensus"
 repeat_region 4095..4188

repeat_region /note="47 copies 2 mer ct 69% conserved"
 4192..4433
 /note="AluSc repeat: matches 135. 309 of consensus"
 repeat_region 4437..6353
 /note="L1M2 repeat: matches 132. 2403 of consensus"
 repeat_region 6521..6942
 /note="L1M2 repeat: matches -707. -278 of consensus"
 repeat_region 7203..9020
 /note="L1P repeat: matches 2297. 4115 of consensus"
 repeat_region 10243..11214
 /note="L1M1 repeat: matches 1555. 2532 of consensus"
 repeat_region 11210..11568
 /note="L1P17 repeat: matches 5749. 6118 of consensus"
 repeat_region 11657..11957
 /note="AluY repeat: matches 1. 299 of consensus"
 repeat_region 11969..12276
 /note="L1M3 repeat: matches -870. -580 of consensus"
 repeat_region 12545..13175
 /note="L2 repeat: matches 2114. 2710 of consensus"
 repeat_region 13304..13609
 /note="AluX repeat: matches 1. 312 of consensus"
 repeat_region 13748..13798
 /note="17 copies 3 mer ttg 80% conserved"
 repeat_region 15340..15435
 /note="MLT1 repeat: matches 318. 410 of consensus"
 repeat_region 16287..16584
 /note="AluSg repeat: matches 1. 306 of consensus"
 repeat_region 16744..17997
 /note="Tigerg3b repeat: matches 4. 1224 of consensus"
 repeat_region 18397..18771
 /note="THEIC repeat: matches 1. 371 of consensus"
 repeat_region 18772..20391
 /note="THEIC-internal repeat: matches 1. 1578 of
 consensus"
 repeat_region 20394..20762
 /note="THEIC repeat: matches 1. 360 of consensus"
 repeat_region 24284..24331
 /note="16 copies 3 mer agy 75% conserved"
 misc_feature 25585..25861
 /note="Sequence from overlapping clone
 BA315A9(AL162498). Assembly confirmed by restriction
 digest."
 repeat_region 27691..27854
 /note="MSTR repeat: matches 1. 177 of consensus"
 repeat_region 27854..27960
 /note="THEIC repeat: matches 316. 422 of consensus"
 misc_feature 28176..28487
 /note="Sequence from overlapping clone
 BA315A9(AL162498). Assembly confirmed by restriction
 digest."
 repeat_region 28418..28850
 /note="L1M2 repeat: matches 5318. 5765 of consensus"
 repeat_region 28883..28924
 /note="21 copies 2 mer ta 76% conserved"
 repeat_region 28945..29194
 /note="L1M2 repeat: matches 4978. 5229 of consensus"
 repeat_region 29235..29844
 /note="L1M2 repeat: matches 4297. 4914 of consensus"
 repeat_region 30076..30415
 /note="L1M2 repeat: matches 2. 344 of consensus"
 misc_feature 30525..30717
 /note="Sequence from overlapping clone
 BA315A9(AL162498). Assembly confirmed by restriction
 digest."
 repeat_region 35507..35598
 /note="AluO/FRAM repeat: matches 157. 249 of consensus"
 repeat_region 35711..35961
 /note="L1M3 repeat: matches 6054. 6304 of consensus"
 repeat_region 36289..36663
 /note="L1M4 repeat: matches 5911. 6299 of consensus"
 repeat_region 37109..37424
 /note="AluSg repeat: matches 3. 308 of consensus"
 repeat_region 38353..39269

```

repeat_region /note="LIMA10 repeat: matches 5409. .6317 of consensus"
39487. .39579
/note="MIR repeat: matches 166. .261 of consensus"
39604. .39737
/note="LIME repeat: matches 5308. .5443 of consensus"
40178. .40217
/note="20 copies 2 mer ct 97% conserved"
40226. .40654
/note="LIM4 repeat: matches 4141. .4604 of consensus"
42776. .42883
/note="54 copies 2 mer aa 63% conserved"
42977. .43059
/note="L2 repeat: matches 2612. .2691 of consensus"
43186. .43309
/note="MLTIE repeat: matches 296. .419 of consensus"
43887. .44226
/note="L2 repeat: matches 1882. .2236 of consensus"
44749. .44859
/note="LIMC3 repeat: matches 6701. .6810 of consensus"
44903. .44935
/note="LIMC4 repeat: matches 6868. .6900 of consensus"
46532. .46978
/note="LTR40b repeat: matches 4. .462 of consensus"
46667. .46847
/note="Sequence from overlapping clone
ba315A9(AL162498). Assembly confirmed by restriction
digest."
46980. .47254
repeat_region /note="HERVL40 repeat: matches 1. .278 of consensus"
48030. .48134
/note="L2 repeat: matches 2015. .2117 of consensus"
48140. .48519
/note="THEIC repeat: matches 1. .369 of consensus"
48530. .48581
/note="26 copies 2 mer aa 71% conserved"
48650. .48700
/note="L2 repeat: matches 2211. .2261 of consensus"
51749. .51964
/note="L2 repeat: matches 1825. .2054 of consensus"
52481. .53602
/note="LTR28 repeat: matches 1. .1735 of consensus"
54284. .54627
/note="LIP12 repeat: matches 5746. .6108 of consensus"
54630. .55318
/note="L1 repeat: matches 4538. .5260 of consensus"
56627. .57090
/note="MLTID repeat: matches 9. .505 of consensus"
57545. .58643
/note="LIMA8 repeat: matches 5148. .6287 of consensus"
58669. .58845
/note="THEIC repeat: matches 1. .177 of consensus"
58846. .59077
/note="LIPAS repeat: matches 5909. .6143 of consensus"
59078. .59332
/note="THEIC repeat: matches 177. .425 of consensus"
59333. .59676
/note="LIMA8 repeat: matches 4787. .5125 of consensus"
59755. .59837
/note="LIMD1 repeat: matches 6133. .6216 of consensus"
59829. .60433
/note="LIMD repeat: matches 83. .708 of consensus"

Query Match 67.9%; Score 19; DB 9; Length 209743;
Best Local Similarity 81.5%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 2 GTCTGGGATAGTCGTCATGCTT 28
DB 181468 GTCTGGGATAGTCGTCATGCTT 181442

```

Search completed: July 10, 2003, 19:20:39
 Job time : 251.625 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003. Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 59.0831 Seconds
(without alignments)
1067.241 Million cell updates/sec

Title: US-10-087-631B-9

Perfect score: 28
Sequence: 1 cgttcgtggagtagctcgtcgtgtgtt 28

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	65.7	510	24	ABO41020
2	18.4	65.7	510	24	ABO41021
3	18.4	65.7	565	24	ABO19830
4	18.4	65.7	565	24	ABO19831
5	18.4	65.7	2636	23	ABL04197
6	18.4	65.7	7330	23	ABL04196
7	18.4	64.3	6265	14	AAQ49460
8	17.6	62.9	34088	23	AAS59566
9	17.4	62.1	297	24	ABN76849

c	10	17.4	62.1	346	21	ABO62757	Mycobacterium tube
c	11	17.4	62.1	541	24	ABO46568	Oligonucleotide fo
c	12	17.4	62.1	541	24	ABO46569	Oligonucleotide fo
c	13	17.4	62.1	654	24	ABN24612	Human ORF polynuc
c	14	17.4	62.1	1149	22	AAE63738	Drosophila gustato
c	15	17.4	62.1	1755	23	ABL27479	Drosophila melanog
c	16	17.4	62.1	2976	22	AAF81351	Quorum sensing con
c	17	17.4	62.1	3836	15	AAO53997	Vitamin D hydroxyl
c	18	17.4	62.1	4337	23	ABL27478	Drosophila melanog
c	19	17.4	62.1	5356	21	AAA46169	GST-GFP fusion pro
c	20	17.4	62.1	6494	21	AAA46170	GFP-Sml-7 fusion p
c	21	17.4	62.1	8549	18	AAV74316	Staphylococcus aur
c	22	17.4	62.1	11460	21	AAZ87211	VEE replicon compr
c	23	17.4	62.1	11795	22	AAH98981	Human EST-derived
c	24	17.4	62.1	11835	23	AAS84466	DNA encoding novel
c	25	17.4	62.1	43095	21	AAA68254	Bacteriophage 3A c
c	26	17.4	62.1	4403765	22	AA199683	Mycobacterium tube
c	27	17.4	62.1	4411529	22	AA199682	DNA encoding novel
c	28	17.2	61.4	462	23	AAS80504	Gene #3877 used to
c	29	17.2	61.4	1247	24	ABN97379	Human polynucleoti
c	30	17.2	61.4	1310	22	AA159616	Human polynucleoti
c	31	17.2	61.4	1320	22	AA157830	Human polynucleoti
c	32	17.2	60.7	135	16	AAV19348	Human gene signal
c	33	17.2	60.7	135	16	AAV19348	Human gene signal
c	34	17.2	60.7	381	21	AAAC07458	Human secreted pro
c	35	17.2	60.7	467	21	AAA28168	Human purH amplico
c	36	17.2	60.7	588	23	ABL11249	Drosophila melanog
c	37	17.2	60.7	725	21	AAZ55634	DNA encoding novel
c	38	17.2	60.7	725	21	AAZ55634	Fennugreek galactom
c	39	17.2	60.7	939	24	ABK68668	Human DNA for olfa
c	40	17.2	60.7	940	24	ABO40952	Oligonucleotide fo
c	41	17.2	60.7	940	24	ABO40953	Oligonucleotide fo
c	42	17.2	60.7	1290	19	AAV53208	Human olfactory OL
c	43	17.2	60.7	1314	21	AAZ55631	Fennugreek galactom
c	44	17.2	60.7	1415	21	AAZ55635	Fennugreek galactom
c	45	17.2	60.7	2696	14	AAO41061	Ap Serotype 7 60KD

ALIGNMENTS

RESULT 1	ABO41020	standard; DNA; 510 BP.
ID	ABO41020	
AC	ABO41020;	
DT	12-JUL-2002 (first entry)	
DE	Oligonucleotide for detecting cytosine methylation SEQ ID NO 27611.	
KW	Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;	
KW	drug; side effect; cancer; central nervous system; cardiovascular;	
KW	gastrintestinal; respiratory system; single nucleotide polymorphism;	
KW	SNP; cell differentiation; ds.	
OS	Homo sapiens.	
PN	WO200218632-A2.	
PD	07-MAR-2002.	
PF	01-SEP-2001; 2001WO-EP10074.	
PR	01-SEP-2000; 2000DE-1043826.	
PR	05-SEP-2000; 2000DE-1044543.	
PA	(EPIC-) EPICENOMICS AG.	
PI	Olek A, Piepenbrock C, Berlin K, Guetig D;	
DR	WPI; 2002-371829/40.	

PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (1) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (11) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

SO Sequence 510 BP; 112 A; 45 C; 151 G; 202 T; 0 other;

Query Match 65.7%; Score 18.4; DB 24; Length 510;
Best Local Similarity 78.6%; Pred. No. 43;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CGTTCGTGGATGATCCGTATGCTT 28
DB 477 TTTT TTTT TTTT TTTT TTTT
CGTTCGTGGATGATGTTAGTGTCTT 504

RESULT 2

ABQ41021/C

ID ABQ41021 standard; DNA; 510 BP.

XX ABQ41021;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 27612.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

XX SNP; cell differentiation; ds.

OS Homo sapiens.

XX WO200218632-A2.

PN 07-MAR-2002.

PD 01-SEP-2001; 2001WO-EP10074.

PF 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIGENOMICS AG.

PA Olek A, Piepenbrock C, Berlin K, Guetig D;

PI WPI; 2002-371829/40.

DR WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (1) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (11) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

SO Sequence 510 BP; 202 A; 151 C; 45 G; 112 T; 0 other;

Query Match 65.7%; Score 18.4; DB 24; Length 510;
Best Local Similarity 78.6%; Pred. No. 43;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CGTTCGTGGATGATCCGTATGCTT 28
DB 34 TTTT TTTT TTTT TTTT TTTT
CGTTCGTGGATGATGTTAGTGTCTT 7

RESULT 3

ABQ19830

ID ABQ19830 standard; DNA; 565 BP.

XX ABQ19830;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 6421.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

XX SNP; cell differentiation; ds.

OS Homo sapiens.

XX WO200218632-A2.

PN 07-MAR-2002.

PD 01-SEP-2001; 2001WO-EP10074.

PF 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIGENOMICS AG.

PA Olek A, Piepenbrock C, Berlin K, Guetig D;

PI WPI; 2002-371829/40.

DR WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert in a
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABO13410-ABO54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX Sequence 565 BP; 123 A; 49 C; 167 G; 226 T; 0 other;

XX Query Match 65.7%; Score 18.4; DB 24; Length 565;

XX Best Local Similarity 78.6%; Pred. No. 44;

XX Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CGTGTGGGATGATCGTCATGCTGT 28

DB 376 CGTGTGGGATGATGCTGTAGTGT 403

RESULT 4

XX ABO19831/C standard; DNA; 565 BP.

XX ABO19831;

XX 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 6422.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

XX drug; side effect; cancer; central nervous system; cardiovascular;

XX gastrointestinal; respiratory system; single nucleotide polymorphism;

XX SNP; cell differentiation; ds.

XX Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI: 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful

XX for diagnosis and prognosis, comprises selective hybridization of

XX amplicons from chemically treated DNA

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of

XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a

XX genomic sample of DNA. The sample is treated chemically to convert

CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABO13410-ABO54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX Sequence 565 BP; 226 A; 167 G; 49 C; 123 T; 0 other;

XX Query Match 65.7%; Score 18.4; DB 24; Length 565;

XX Best Local Similarity 78.6%; Pred. No. 44;

XX Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CGTGTGGGATGATCGTCATGCTGT 28

DB 190 CGTGTGGGATGATGCTGTAGTGT 163

RESULT 5

XX ABL04197/C standard; cDNA; 2636 BP.

XX ABL04197;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 7073.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EM;

XX WPI: 2001-656860/75.

XX P-PSDB; ABB60094.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions

XX Claim 1; SEQ ID NO 7073; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX sequences (ABL01840-ABL16175) and the encoded proteins

PN WO200181581-A2.
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
XX PR 02-JUN-2000; 2000US-208841P.
XX PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX WPI; 2001-616774/71.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris -
XX
XX Claim 1; SEQ ID No 61; 1069pp; English.
XX
XX Sequences AAS59506-AAS59804 represent DNA molecules encoding
XX CC propionibacterium acnes immunogenic polypeptides. The proteins and their
XX CC associated DNA sequences are used in the treatment, prevention and
XX CC diagnosis of medical conditions caused by P. acnes. The disorders include
XX CC SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and
XX CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved
XX CC in infections of bone, joints and the central nervous system, however it
XX CC is particularly involved in the inflammatory lesions associated with acne
XX CC vulgaris. A method for detecting the presence or absence of P. acnes in a
XX CC patient comprises contacting a sample with a binding agent that binds to
XX CC the proteins of the invention and determining the amount of bound protein
XX CC in the sample. The polypeptides may be used as antigens in the production
XX CC of antibodies specific for P. acnes proteins. These antibodies can be
XX CC used to downregulate expression and activity of P. acnes polypeptides and
XX CC therefore treat P. acnes infections. The antibodies may also be used as
XX CC diagnostic agents for determining P. acnes presence, for example, by
XX CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
XX CC polypeptides shown in AAV54556-AAV54928 and AAV67560-AAV67562.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 34088 BP; 6991 A; 11002 C; 9453 G; 6639 T; 3 other;
XX
XX Query Match 62.9%; Score 17.6; DB 23; Length 34088;
XX Best Local Similarity 83.3%; Pred. No. 1.9e+02;
XX Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX 1 CGTGTGGATAGTCGCGCATGG 24
XX DB 22579 CGTGTGGATAGTCGCGCATGG 22556
XX
XX RESULT 9
XX ID ABA76849 standard; CDNA; 297 BP.
XX
XX ABA76849;
XX
XX 08-JUL-2002 (first entry).
XX
XX Human ORF1796 cDNA, SEQ ID NO:3591.
XX
XX Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
XX KM disease monitoring; cytokine; cell proliferation; cell differentiation;
XX KM immune modulation; haematopoiesis regulation; tissue growth;
XX KM angiogenesis; activin; inhibitor; chemotactic; chemokine; haemostatic;
XX KM thrombolytic; tumour inhibition; bodily characteristics; fertility;
XX KM behaviour; cancer; proliferative disorder; neurological disorder;
XX KM cardiovascular disease; immune system disorder; organ transplantation;

KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulvarey;
KW vasotropic; antipsoriatic; antidiabetic; cytosolic; neurotic;
KW neuroprotective; antithrombotic; anticoagulant; thrombolytic;
KW cardiant; hypotensive; antihypertensive; antineoplastic; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200190366-A2.
XX
XX 29-NOV-2001.
XX
XX 24-MAY-2001; 2001WO-US17076.
XX
XX 24-MAY-2000; 2000US-206690P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Leach MD, Shimkets RA;
XX
XX WPI; 2002-106200/14.
XX DR P-PSDB; ABP32823.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
XX PT preventing and treating cardiovascular disease, neurodegenerative,
XX PT hyperproliferative disorders and disorders related to organ
XX PT transplantation
XX
XX Claim 1; Page 1157; 2508pp; English.
XX
XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins
XX CC designated ORF (open reading frame) 1-4534, and sequences ABA75054-
XX CC ABA79587 represent cDNAs encoding them. The invention also encompasses
XX CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
XX CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
XX CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
XX CC polynucleotides, the recombinant production of ORFX proteins, antibodies
XX CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
XX CC polypeptides, methods of screening for modulators of ORFX expression or
XX CC activity, and methods of screening individuals for a predisposition to an
XX CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
XX CC range of biological activities, such as cytokine, cell proliferation,
XX CC cell differentiation, immune modulation, haematopoiesis regulation,
XX CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
XX CC chemokinetic activity, haemostatic activity, thrombolytic activity,
XX CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
XX CC and antiinfective activity, and may also be involved in the determination
XX CC of bodily characteristics, fertility and behaviour. ORFX proteins,
XX CC nucleic acids and antibodies may be used in the treatment of cancers,
XX CC other proliferative disorders such as psoriasis and benign tumours,
XX CC neurological disorders such as epilepsy and Alzheimer's disease,
XX CC cardiovascular diseases, immune system disorders, disorders related to
XX CC organ transplantation, disorders of tissue growth and regeneration,
XX CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol
XX CC storage disease, and infectious diseases caused by viral, bacterial,
XX CC fungal and other pathogens. ORFX nucleic acids may also be used as a
XX CC source of primers and probes, in the detection of ORFX genomic sequences
XX CC or transcripts, in the identification and cloning of homologous
XX CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
XX CC nucleic acids may additionally be used to produce transgenic animals
XX CC which may be useful for studying the function and/or activity of ORFX
XX CC protein, and in drug screening. The ORFX proteins may also be used as
XX CC immunogens to generate specific antibodies, which are useful in the
XX CC diagnosis, treatment and monitoring of ORFX-associated diseases.
XX
XX Sequence 297 BP; 66 A; 99 C; 78 G; 54 T; 0 other;
XX
XX Query Match 62.1%; Score 17.4; DB 24; Length 297;
XX Best Local Similarity 77.8%; Pred. No. 1.2e+02;
XX Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
XX
XX 2 GTTCTGGATAGTCGCGCATGGT 28

Db 123 GTTCCAGGTCAGTTCGTCATGCTTT 97

RESULT 10

ID AB062757/c standard; DNA; 346 BP.

AC AB062757;

DE 16-AUG-2002: (first entry)

DE Mycobacterium tuberculosis BAC vector clone RV240SP6.

KW Mycobacterium tuberculosis; Mycobacterium bovis; mycobacterium;
KM detection; BAC vector; bacterial artificial chromosome; tuberculosis;
KW gene; ds.

OS Mycobacterium tuberculosis.

PN WO9554487-A2.

PD 28-OCT-1999.

PF 16-APR-1999; 99WO-IB00740.

PR 16-APR-1998; 98US-0060756.

PA (INSP) INST PASTEUR.

PI Cole S, Buchrieser-Brosch R, Gordon S, Billault A;

DR WPI; 2000-013262/01.

PT Isolation of polynucleotides from mycobacterial genomes, useful for
PS detection of Mycobacteria and for combating tuberculosis -

Claim 23: Page 70; 161pp; English.

XX The present invention describes a method for isolating a polynucleotide
CC of interest that is present or is expressed in a genome of a first
CC mycobacterium strain and that is absent or altered in a genome of a
CC second mycobacterium strain, which is different from the first strain
CC using a bacterial artificial chromosome (BAC) vector. Recombinant BAC
CC vectors, which are preferably immobilised, can be used to detect
CC mycobacterial nucleic acids (genomic DNA, cDNA or mRNA) in biological
CC samples. The polynucleotides identified are useful as probes or primers
CC for detecting a given mycobacterium of interest. By aligning the
CC polynucleotides contained in the recombinant BAC vectors it is possible
CC to physically map a polynucleotide of mycobacterial origin in a
CC biological sample. The methods and vectors from the present invention
CC are useful in providing information for combating tuberculosis. It is
CC possible to compare genomes between different strains or species and
CC their non-pathogenic strains or species counterparts. AB062492 to
CC AB063228 and AB881227 to AB881230 represent sequences used in the
CC exemplification of the present invention.

SO Sequence 346 BP; 53 A; 96 C; 125 G; 70 T; 2 other;

Query Match Best Local Similarity 62.1%; Score 17.4; DB 21; Length 346;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GTTCTGGGATGATCGCTC 20

Db 253 GTTCTGGGATGATCGCTC 235

RESULT 11

ID AB046568 standard; DNA; 541 BP.

AC AB046568;

XX 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 33159.

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KM drug; side effect; cancer; central nervous system; cardiovascular;
KM gastrointestinal; respiratory system; single nucleotide polymorphism;
KM SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guelzig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful
PS for diagnosis and prognosis, comprises selective hybridization of
PS amplicons from chemically treated DNA

Claim 12: 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP/s); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

SO Sequence 541 BP; 82 A; 59 C; 201 G; 197 T; 2 other;

Query Match Best Local Similarity 62.1%; Score 17.4; DB 24; Length 541;

Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GTTCTGGGATGATCGCTC 28

Db 450 GTTCTGGGATGATCGCTC 476

RESULT 12

ID AB046569/c standard; DNA; 541 BP.

AC AB046569;

DT 12-JUL-2002 (first entry)

OS	Homo sapiens..
XX	
PN	W0200192523-A2.
XX	
PD	06-DEC-2001.
PF	29-MAY-2001; 2001WO-US10836.
XX	
PR	30-MAY-2000; 2000US-206132P.
PR	29-AUG-2000; 2000US-228716P.
PA	(CURA-) CURAGEN CORP.
PI	Shimkets RA, Leach MD;
XX	
DR	WPI: 2002-106308/14.
XX	P-PSDB; ABP08860.
PT	
PT	Novel human polypeptides and polynucleotides useful for diagnosing,
PS	hyperproliferative disorders and autoimmune disorders
XX	
XX	Disclosure; SEQ ID 17701; 1037pp; English.
CC	
CC	The present invention describes substantially purified human proteins
CC	(referred to as open reading frame, ORFX, where x is 1-11491 (see Table 1
CC	in the specification)). ABN15762 to ABN27252 encode the human ORFX
CC	proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC	treating or preventing a pathology associated with an ORFX-associated
CC	disorder in humans, and in the manufacture of a medicament for treating a
CC	syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC	sequences can be used in gene therapy. ORFX sequences can be used in the
CC	treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC	psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC	osteoarthritis, neurodegenerative disorders, disorders related to organ
CC	transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC	lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC	storage disease, various immune deficiencies and disorders, infectious
CC	diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC	arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC	disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC	useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC	bone degenerative disorders, or periodontal disease, and for gut
CC	protection or regeneration and treatment of lung or liver fibrosis,
CC	repertusion injury in various tissues and conditions resulting from
CC	systemic cytokine damage.
CC	N.B. The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from Wipo
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 654 BP; 90 A; 194 C; 180 G; 190 T; 0 other:
Query Match	62.1%; Score-17.4; DB 24; Length 654;
Best Local Similarity	77.8%; Pmid.No. 1.3e+02;
Matches 21; Conservative	0; Mismatches 6; Indels 0; Gaps 0;
OY	1 CGTTCGGGGAATAGTCGGTATCGGTCT 27
Db	
	344 CGTTCGGGGAATAGTCGGCAGGGGTCT 370
RESULT 14	
AAP63738	
ID	AAP63738 standard; DNA; 1149 BP.
XX	

AC AAF63738;
 XX
 DT 03-APR-2001 (first entry)
 XX
 DE Drosophila gustatory receptor GR23A.1a DNA sequence.
 XX
 KM Gustatory receptor; fruit fly; taste; pheromone; semiochemical;
 XX crop damage; pest control; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200077208-A2.
 XX
 PD 21-DEC-2000.
 XX
 PF 14-JUN-2000; 2000WO-US16211.
 XX
 PR 14-JUN-1999; 99US-0138668.
 XX
 PR 10-FEB-2000; 2000US-0181704.
 XX
 PA (UYVA) UNIV YALE.
 XX
 PI Carlson PJ, Clyne PJ, Mair CG;
 XX
 DR WPI; 2001-061873/07.
 XX
 DR P-PSDB: AAB75199.
 XX
 PT New isolated nucleic acid molecule encoding Drosophila gustatory
 PT Receptor protein useful for e.g. identification of compounds which may
 PT be used for pest management -
 XX
 PS Claim 3; Page 95-97; 227pp. English.
 XX
 CC This invention relates to polynucleotide sequences AAF63732 - AAF63777
 CC which encode Drosophila gustatory receptor proteins represented by
 CC sequences AAB75193 - AAB75238. The invention includes methods for
 CC determining gustatory receptor ligands. Also included is a method for
 CC modulating the expression of the DNA encoding the receptors. The DNA and
 CC protein sequences may be used for the identification of compounds,
 CC e.g. pheromones and other semiochemicals, which may be used for pest
 CC management. The DNA sequences may also be used for behavioural studies
 CC involving gustatory systems in various organisms. Also, the DNA sequences
 CC may also be used to track down gustatory receptor genes in insects that
 CC damage crops or transmit diseases.
 XX
 SQ Sequence 1149 BP; 255 A; 276 C; 292 G; 326 T; 0 other;
 Query Match 62.1%; Score 17.4; DB 22; Length 1149;
 Best Local Similarity 77.8%; Pred. NO. 1.4e+02;
 Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 OY 2 GTTCGTGGGATAGTCGTCATGCGTCTT 28
 DB 258 GTTCGTGGGAAACTCGTCTGGGATT 284
 RESULT 15
 ABL27479
 ID ABL27479 standard; DNA; 1755 BP.
 XX
 AC ABL27479;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 33910.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical; gene; ds.
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX

PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EM;
 XX
 DR WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 33910; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins
 CC (AAB57737-AB12072).
 CC
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 1755 BP; 386 A; 425 C; 437 G; 507 T; 0 other;
 Query Match 62.1%; Score 17.4; DB 23; Length 1755;
 Best Local Similarity 77.8%; Pred. NO. 1.5e+02;
 Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 OY 2 GTTCGTGGGATAGTCGTCATGCGTCTT 28
 DB 219 GTTCGTGGGAAACTCGTCTGGGATT 245

Search completed: July 10, 2003, 19:52:53
 Job time : 67.0831 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 ; Search time 12.5213 Seconds
(without alignments)
685.785 Million cell updates/sec

Title: US-10-087-631B-9

Perfect score: 28

Sequence: 1 cgttcgtggtatgctcgtcgtgtgtt 28

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2.6/ptodata/1/ina/5A.COMB.seq:**

2: /cgn2.6/ptodata/1/ina/5B.COMB.seq:**

3: /cgn2.6/ptodata/1/ina/6A.COMB.seq:**

4: /cgn2.6/ptodata/1/ina/6B.COMB.seq:**

5: /cgn2.6/ptodata/1/ina/PC10S.COMB.seq:**

6: /cgn2.6/ptodata/1/ina/Backfile1.seq:**

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.6	62.9	2476	US-09-221-017B-876	Sequence 876, App
2	17.4	62.1	346	US-09-060-756-256	Sequence 256, App
3	17.4	62.1	4403765	US-09-103-840A-2	Sequence 1, Appl
4	17.4	62.1	4411529	US-09-103-840A-1	Sequence 1, Appl
5	17	60.7	1290	US-08-827-291A-1	Sequence 1, Appl
6	17	60.7	2696	US-07-961-522-1	Sequence 1, Appl
7	17	60.7	2696	US-08-217-438-1	Sequence 1, Appl
8	17	60.7	2696	US-08-321-978-1	Sequence 1, Appl
9	17	60.7	2696	US-08-710-584-1	Sequence 1, Appl
10	16.8	60.0	2418	US-09-388-743-25	Sequence 25, Appl
11	16.8	60.0	2781	US-09-342-681C-116	Sequence 116, App
12	16.8	60.0	4235	US-09-342-681C-18	Sequence 18, App
13	16.8	60.0	5476	US-09-221-017B-147	Sequence 147, App
14	16.6	59.3	367	US-08-936-165A-217	Sequence 217, App
15	16.6	59.3	948	US-09-134-001C-724	Sequence 724, App
16	16.4	58.6	405	US-08-806-597A-7	Sequence 7, Appl
17	16.4	58.6	405	US-08-970-428A-7	Sequence 7, Appl
18	16.4	58.6	1050	US-08-806-597A-13	Sequence 13, Appl
19	16.4	58.6	1050	US-08-970-428A-13	Sequence 13, Appl
20	16.4	58.6	2588	PCR-US91-02954-1	Sequence 1, Appl
21	16.4	58.6	2608	US-07-725-083-1	Sequence 1, Appl
22	16.2	57.9	1817	US-08-538-816A-3	Sequence 3, Appl
23	16.2	57.9	1817	US-09-076-651-3	Sequence 3, Appl
24	16.2	57.9	1817	US-09-208-394-3	Sequence 3, Appl
25	16.2	57.9	1844	US-08-538-816A-10	Sequence 10, Appl
26	16.2	57.9	1844	US-09-076-651-10	Sequence 10, Appl
27	16.2	57.9	1844	US-09-208-394-10	Sequence 10, Appl

28	16.2	57.9	1894	US-08-538-816A-8	Sequence 8, Appl
29	16.2	57.9	1894	US-09-076-651-8	Sequence 8, Appl
30	16.2	57.9	1894	US-09-208-394-8	Sequence 8, Appl
31	16	57.1	1799	US-08-687-590-58	Sequence 58, Appl
32	16	57.1	1988	US-08-961-527-268	Sequence 268, App
33	16	57.1	3073	US-08-975-762-41	Sequence 41, Appl
34	16	57.1	3073	US-09-295-028-41	Sequence 41, Appl
35	16	57.1	3073	US-09-106-582-41	Sequence 41, Appl
36	16	57.1	6122	US-08-403-545-1	Sequence 1, Appl
37	16	57.1	6122	US-08-404-381-1	Sequence 1, Appl
38	16	57.1	7032	US-08-149-097D-24	Sequence 24, Appl
39	16	57.1	7032	US-08-949-386-24	Sequence 24, Appl
40	16	57.1	7032	US-08-450-562-24	Sequence 24, Appl
41	16	57.1	7032	US-08-984-709A-24	Sequence 24, Appl
42	16	57.1	7032	US-08-450-272-24	Sequence 24, Appl
43	16	57.1	7089	US-08-949-386-25	Sequence 25, Appl
44	16	57.1	7089	US-08-450-562-25	Sequence 25, Appl
45	16	57.1	7089	US-08-984-709A-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-09-221-017B-876/C
Sequence 876, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: ROSS, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Montoy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 876:
SEQUENCE CHARACTERISTICS:
LENGTH: 2476 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..2476
US-09-221-017B-876

Query Match
Best Local Similarity 62.1%; Score 17.6; DB 4; Length 2476;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db
5 CGTGGATAGTCGTCATGCTGT 28
996 CGTGGATGTCGTCACGAGTT 973

RESULT 2

US-09-060-756-256/C
Sequence 256, Application US/09060756
Patent No. 6183957

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: Cole, Stewart
APPLICANT: Buchleser-Brosch, Roland
APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain

TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM

TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA

FILE REFERENCE: 3495-0169

CURRENT APPLICATION NUMBER: US/09/060,756

NUMBER OF SEQ ID NOS: 743

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 256

LENGTH: 346

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

FEATURE:

NAME/KEY: unsure

LOCATION: (various positions within the sequence)

OTHER INFORMATION: Applicants are uncertain of bases designated as "n".

US-09-060-756-256

Query Match
Best Local Similarity 62.1%; Score 17.4; DB 4; Length 346;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db
2 GTTCGTGGATAGTCGTCGTC 20
253 GTTCGTGGATAGTCGTCGTC 235

RESULT 3

US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2
LENGTH: 4403765

TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match
Best Local Similarity 62.1%; Score 17.4; DB 4; Length 4403765;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db
1 CGTTCGTGGATAGTCGTCATGCTGT 27
3836409 CGTTCGTGGATAGTCGTCATGCTGT 3836435

RESULT 4

US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 4411529

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

OTHER INFORMATION: H37Rv

US-09-103-840A-1

Query Match
Best Local Similarity 62.1%; Score 17.4; DB 4; Length 4411529;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db
1 CGTTCGTGGATAGTCGTCATGCTGT 27
3850511 CGTTCGTGGATAGTCGTCATGCTGT 3850537

RESULT 5

US-08-827-291A-1
Sequence 1, Application US/08827291A
Patent No. 5874243

GENERAL INFORMATION:

APPLICANT: Macina, Roberto
APPLICANT: Sathe, Ganesh

TITLE OF INVENTION: NOVEL OLRC15 RECEPTOR

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: 19406

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0

APPLICATION NUMBER: US/08/827,291A
FILING DATE: 28-MAR-1997

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: KING, WILLIAM T
REGISTRATION NUMBER: 30,954
REFERENCE/DOCKET NUMBER: GP50001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5015
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1290 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-827-291A-1

Query Match 60.7%; Score 17; DB 2; Length 1290;
Best Local Similarity 80.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 CGTTCGTGATAGTCCGTCATGCT 25
|||||
DB 408 CCTCATGGGAACCTCTCATGG 432

RESULT 6
US-07-961-522-1
Sequence 1, Application US/07961522
Patent No. 5417971
GENERAL INFORMATION:
APPLICANT: POTTER, ANDREW A.
APPLICANT: GERLACH, GERALD F.
APPLICANT: WILSON, PHILIP J.
APPLICANT: ROSSI-CAMPOS, AMALIA
TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
TITLE OF INVENTION: PLEUROSPERMIONIAE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROBERTA L. ROBINS
STREET: 635 BRYANT STREET
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/961,522
FILING DATE: 19921015
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9000-0015.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 617-8999
TELEFAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2696 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:

NAME/KEY: CDS
LOCATION: 333..1973
US-07-961-522-1

Query Match 60.7%; Score 17; DB 1; Length 2696;
Best Local Similarity 80.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 TTCGTGGATAGTCCGTCATGCTGT 27
|||||
DB 2267 TTCGTGGGATAGTCAATGCTGT 2291

RESULT 7
US-08-217-438-1
Sequence 1, Application US/08217438
Patent No. 5521072
GENERAL INFORMATION:
APPLICANT: Potter, Andrew A.
APPLICANT: Gerlach, Gerald F.
APPLICANT: Wilson, Philip J.
APPLICANT: Rossi-Campos, Amalia
TITLE OF INVENTION: ACTINOBACILLUS PLEUROSPERMIONIAE
TITLE OF INVENTION: TRANSFERRIN BINDING PROTEINS AND USES THEREOF
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Reed & Robins
STREET: 635 Bryant Street
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/217,438
FILING DATE: 22-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Robins, Roberta L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9001-0015.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-617-8999
TELEFAX: 415-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2696 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1973
US-08-217-438-1

Query Match 60.7%; Score 17; DB 1; Length 2696;
Best Local Similarity 80.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 TTCGTGGATAGTCCGTCATGCTGT 27
|||||
DB 2267 TTCGTGGGATAGTCAATGCTGT 2291

RESULT 8
US-08-321-978-1
Sequence 1, Application US/08321978
Patent No. 5801018

GENERAL INFORMATION:
APPLICANT: POTTER, ANDREW A.
APPLICANT: GERLACH, GERALD F.
APPLICANT: WILSON, PHILIP J.
APPLICANT: ROSSI-CAMPOS, AMALIA
TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROBERTA L. ROBINS
STREET: 285 HAMILTON AVE, SUITE 200
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/321,978
FILING DATE: 12-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9001-0015.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 327-3400
FAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2696 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1973
US-08-321-978-1

Query Match 60.7% Score 17; DB 1; Length 2696;
Best Local Similarity 80.0% Pred. No. 36;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 TTCGTGGATAGTCGTCATGATGTT 27
Db 2267 TTCGTGGGATAGTCGTAATCGTGT 2291

RESULT 9
US-08-710-584-1
Sequence 1, Application US/08710584
Patent No. 5876725
GENERAL INFORMATION:
APPLICANT: POTTER, ANDREW A.
APPLICANT: GERLACH, GERALD F.
APPLICANT: WILSON, PHILIP J.
APPLICANT: ROSSI-CAMPOS, AMALIA
TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROBERTA L. ROBINS
STREET: 285 HAMILTON AVE, SUITE 200
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/710,584
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/321,978
FILING DATE: 12-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9001-0015.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 327-3400
FAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2696 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1973
US-08-710-584-1

Query Match 60.7% Score 17; DB 2; Length 2696;
Best Local Similarity 80.0% Pred. No. 36;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 TTCGTGGATAGTCGTCATGATGTT 27
Db 2267 TTCGTGGGATAGTCGTAATCGTGT 2291

RESULT 10
US-09-388-743-25
Sequence 25, Application US/09388743
Patent No. 6423886
GENERAL INFORMATION:
APPLICANT: Singleary, George
APPLICANT: Zhou, Ian
TITLE OF INVENTION: Use in the Production of New Starches
FILE REFERENCE: 1144
CURRENT APPLICATION NUMBER: US/09/388,743
CURRENT FILING DATE: 1999-09-02
NUMBER OF SEQ ID NOS: 28
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 25
LENGTH: 2418
TYPE: DNA
ORGANISM: Typha latifolia
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(2418)
US-09-388-743-25

Query Match 60.0% Score 16.8; DB 4; Length 2418;
Best Local Similarity 75.0% Pred. No. 44;
Matches 21; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CTTGCTGGATAGTCGTCATGATGTT 28
Db 1656 CATGCCCTGATGTGTCAGTCATGATGTT 1683

RESULT 11
US-09-342-681C-116/C
Sequence 116, Application US/09342681C

Patent No. 6355782
GENERAL INFORMATION:
APPLICANT: Zonana et al.
TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
FILE REFERENCE: 52978
CURRENT APPLICATION NUMBER: US/09/342,681C
CURRENT FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: 60/092,279
PRIOR FILING DATE: 1998-07-09
PRIOR APPLICATION NUMBER: 60/112,366
PRIOR FILING DATE: 1998-12-15
NUMBER OF SEQ ID NOS: 123
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 116
LENGTH: 2781
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (528)..(2756)
OTHER INFORMATION: n l s a, c, t o r g
US-09-342-681C-116

Query Match 60.0%; Score 16.8; DB 4; Length 2781;
Best Local Similarity 75.0%; Pred. No. 45;
Matches 21; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 CGTTCGTGGAGTACGCTCATGCTT 28
DB 2010 CGTTCGTGGAGTACGCTCATGCTT 1983

RESULT 12
US-09-342-681C-18/C
Sequence 18, Application US/09342681C
Patent No. 6355782
GENERAL INFORMATION:
APPLICANT: Zonana et al.
TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
FILE REFERENCE: 52978
CURRENT APPLICATION NUMBER: US/09/342,681C
CURRENT FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: 60/092,279
PRIOR FILING DATE: 1998-07-09
PRIOR APPLICATION NUMBER: 60/112,366
PRIOR FILING DATE: 1998-12-15
NUMBER OF SEQ ID NOS: 123
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 18
LENGTH: 4235
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (433)..(1779)
US-09-342-681C-18

Query Match 60.0%; Score 16.8; DB 4; Length 4235;
Best Local Similarity 75.0%; Pred. No. 48;
Matches 21; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 CGTTCGTGGAGTACGCTCATGCTT 28
DB 3466 CGTTCGTGGAGTACGCTCATGCTT 3439

RESULT 13
US-09-221-017B-147
Sequence 147, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM type: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H.
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 147:
SEQUENCE CHARACTERISTICS:
LENGTH: 5476 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...5476
US-09-221-017B-147

Query Match 60.0%; Score 16.8; DB 4; Length 5476;
Best Local Similarity 75.0%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 CGTTCGTGGAGTACGCTCATGCTT 28
DB 5013 CATTCGTGGAGAGCCCGCATGCTGAT 5040

RESULT 14
US-08-936-165A-217
Sequence 217, Application US/08936165A
Patent No. 6348582
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Burnham, Martin
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Lonetto, Michael
APPLICANT: Nicholas, Richard

us-10-087-631b-9.rn1

Page 6

```

Query Match          59.3%; Score 16.6; DB 4; Length 948;
Best Local Similarity 82.6%; Pred. No 47;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

0Y      1 CGTCTGTCGGAGATGTCCTCATG 23
          |||||  |||||  |  |||||
Db      106 CGTCTGTCGATGATGCTCATG 84

Search completed: July 10, 2003, 20:28:02
Job time : 33.5213 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:53:16 ; Search time 76.6382 Seconds

(without alignments)
576.891 Million cell updates/sec

Title: US-10-087-631B-9

Perfect score: 28

Sequence: 1 cgttcgtggatagtcgctgctgtgtt 28

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications, NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	28	100.0	28	US-10-087-631B-9	Sequence 9, Appl1
2	28	100.0	241	US-10-087-631B-12	Sequence 12, Appl1
3	28	100.0	943	US-10-087-631B-11	Sequence 11, Appl1
4	20	71.4	1533	US-10-156-761-6429	Sequence 6429, Ap
5	20	71.4	9025608	US-10-156-761-1	Sequence 1, Appl1
6	17.4	62.1	1572	US-10-156-761-5471	Sequence 5471, Ap
7	17.4	62.1	8549	US-08-781-986A-5	Sequence 5, Appl1
8	17.4	62.1	684973	US-09-263-959-1	Sequence 1, Appl1
9	17.4	62.1	9025608	US-10-156-761-1	Sequence 1, Appl1
10	17.2	61.4	1247	US-09-880-107-3874	Sequence 3874, Ap
11	17.2	61.4	1320	US-10-098-841-33	Sequence 33, Appl1
12	17	60.7	493	US-09-796-692-4625	Sequence 4625, Ap
13	17	60.7	493	US-10-040-862-4625	Sequence 4625, Ap
14	17	60.7	924	US-10-156-761-1019	Sequence 1019, Ap
15	17	60.7	1290	US-09-826-508-23	Sequence 23, Appl1
16	16.8	60.0	1338	US-09-815-242-7956	Sequence 7956, Ap
17	16.8	60.0	2781	US-09-729-658B-116	Sequence 116, Ap
18	16.8	60.0	3623	US-09-918-909-23	Sequence 23, Appl1
19	16.8	60.0	4235	US-09-729-658B-18	Sequence 18, Appl1

20	16.6	59.3	367	10	US-09-939-980-217	Sequence 217, App
21	16.6	59.3	567	10	US-09-864-761-16516	Sequence 16516, A
22	16.4	58.6	428	10	US-09-954-456-1334	Sequence 1334, Ap
23	16.4	58.6	583	9	US-10-106-688-3533	Sequence 3533, Ap
24	16.4	58.6	633	12	US-10-029-079-4	Sequence 4, Appl1
25	16.4	58.6	896	9	US-10-007-280A-125	Sequence 125, App
26	16.4	58.6	998	9	US-10-007-280A-126	Sequence 126, App
27	16.4	58.6	1017	9	US-10-198-846-11216	Sequence 11216, A
28	16.4	58.6	1038	10	US-09-974-300-439	Sequence 439, App
29	16.4	58.6	1131	10	US-09-925-302-26	Sequence 26, Appl1
30	16.4	58.6	1325	9	US-08-925-299-185	Sequence 185, App
31	16.4	58.6	1325	10	US-09-925-299-185	Sequence 185, App
32	16.4	58.6	1328	9	US-10-082-830-23	Sequence 23, Appl1
33	16.4	58.6	1467	9	US-10-156-761-1138	Sequence 1138, Ap
34	16.4	58.6	1638	9	US-10-270-333-1170	Sequence 1170, App
35	16.4	58.6	1830	10	US-09-969-147-212	Sequence 212, App
36	16.4	58.6	1830	10	US-09-880-107-3797	Sequence 3797, App
37	16.4	58.6	1979	10	US-09-925-300-224	Sequence 224, App
38	16.4	58.6	2182	9	US-10-091-572-816	Sequence 816, App
39	16.4	58.6	2182	9	US-09-764-891-9240	Sequence 9240, Ap
40	16.4	58.6	2244	7	US-08-781-986A-270	Sequence 270, App
41	16.4	58.6	2723	9	US-10-177-293-372	Sequence 372, App
42	16.4	58.6	2921	10	US-09-925-301-223	Sequence 223, App
43	16.4	58.6	3395	9	US-10-198-846-13880	Sequence 13880, A
44	16.4	58.6	3809	9	US-10-273-680-1	Sequence 1, Appl1
45	16.4	58.6	3957	9	US-10-270-333-169	Sequence 169, App

ALIGNMENTS

RESULT 1

US-10-087-631B-9

Sequence 9, Application US/10087631B

Publication No. US20030054372A1

GENERAL INFORMATION:

APPLICANT: JAEGER, STEPHAN

TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A

TITLE OF INVENTION: CONTROL

FILE REFERENCE: 1803-335-999

CURRENT APPLICATION NUMBER: US/10/087,631B

CURRENT FILING DATE: 2002-03-01

NUMBER OF SEQ ID NOS: 17

SOFTWARE: PatentIn version 3.1

SEQ ID NO: 9

LENGTH: 28

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: 5778pc primer parallel-

US-10-087-631B-9

Query Match 100.0%; Score 28; DB 9; Length 28;

Best Local Similarity 100.0%; Pred. No. 0.00087;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 CGTTCGTGGGATAGTCGCTCATGTGTT 28

US-10-087-631B-12/c

Sequence 12, Application US/10087631B

Publication No. US20030054372A1

GENERAL INFORMATION:

APPLICANT: JAEGER, STEPHAN

TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A

TITLE OF INVENTION: CONTROL

FILE REFERENCE: 1803-335-999

CURRENT APPLICATION NUMBER: US/10/087,631B

CURRENT FILING DATE: 2002-03-01

NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 12
LENGTH: 241
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Amplicon derived from QS (pc
OTHER INFORMATION: using the primers ST280pc and ST778pc
US-10-087-631B-12

Query Match
Best Local Similarity 100.0%; Score 28; DB 9; Length 241;
Pred. No. 0.001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGTTCGTGGATAGTCCGTCATGCTGT 28
|||||
DB 241 CGTTCGTGGATAGTCCGTCATGCTGT 214

RESULT 3
US-10-087-631B-11/c
Sequence 11, Application US/10087631B
Publication No. US20030054372A1
GENERAL INFORMATION:
APPLICANT: JAEGER, STEPHAN
TITLE OF INVENTION: A METHOD FOR THE DETERMINATION OF A NUCLEIC ACID USING A
FILE REFERENCE: 1803-335-999
CURRENT APPLICATION NUMBER: US/10/087,631B
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 11
LENGTH: 943
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: QS (pc) HCV being parallel-
US-10-087-631B-11

Query Match
Best Local Similarity 100.0%; Score 28; DB 9; Length 943;
Pred. No. 0.0011;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGTTCGTGGATAGTCCGTCATGCTGT 28
|||||
DB 297 CGTTCGTGGATAGTCCGTCATGCTGT 270

RESULT 4
US-10-156-761-6429/c
Sequence 6429, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 6429

LENGTH: 1533
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1533)
US-10-156-761-6429

Query Match
Best Local Similarity 71.4%; Score 20; DB 9; Length 1533;
Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGTTCGTGGATAGTCCGTCATGCTGT 28
|||||
DB 196 CGATCGTGTGACGTCCTGTGCTGT 169

RESULT 5
US-10-156-761-1
Sequence 1, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match
Best Local Similarity 71.4%; Score 20; DB 9; Length 9025608;
Pred. No. 10;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGTTCGTGGATAGTCCGTCATGCTGT 28
|||||
DB 7731103 CGATCGTGTGACGTCCTGTGCTGT 7731130

RESULT 6
US-10-156-761-5471
Sequence 5471, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 5471
LENGTH: 1572
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1572)
US-10-156-761-5471

Query Match 62.1%; Score 17.4; DB 9; Length 1572;
Best Local Similarity 77.8%; Pred. No. 99;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GTTCGTGGGATAGTCGTCATGCTGT 28
DB 1143 GTTCGTGGGATAGTCGTCATGCTGT 1169

RESULT 7
US-08-781-986A-5/C
Sequence 5, Application US/08781986A
Publication No. US20030054436A1

GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8549 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-5

Query Match 62.1%; Score 17.4; DB 7; Length 8549;
Best Local Similarity 77.8%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GTTCGTGGGATAGTCGTCATGCTGT 28
DB 6487 GTTCATAGGCTAGTCTGTAATAGTGT 6461

RESULT 8
US-09-263-959-1

Sequence 1, Application US/09263959
Patent No. US20020150891A1

GENERAL INFORMATION:
APPLICANT: Hood, Leroy E.
APPLICANT: Rowen, Lee
APPLICANT: Koop, Ben F.
TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH U
NUMBER OF SEQUENCES: 1279
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/263,959
FILING DATE: 05-MAR-1999
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: McMaisters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 920010.426C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 684973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-263-959-1

Query Match 62.1%; Score 17.4; DB 10; Length 684973;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GTTCGTGGGATAGTCGTCATGCTGT 28
DB 513413 GTTCGTGGGATAGTCGTCATGCTGT 513439

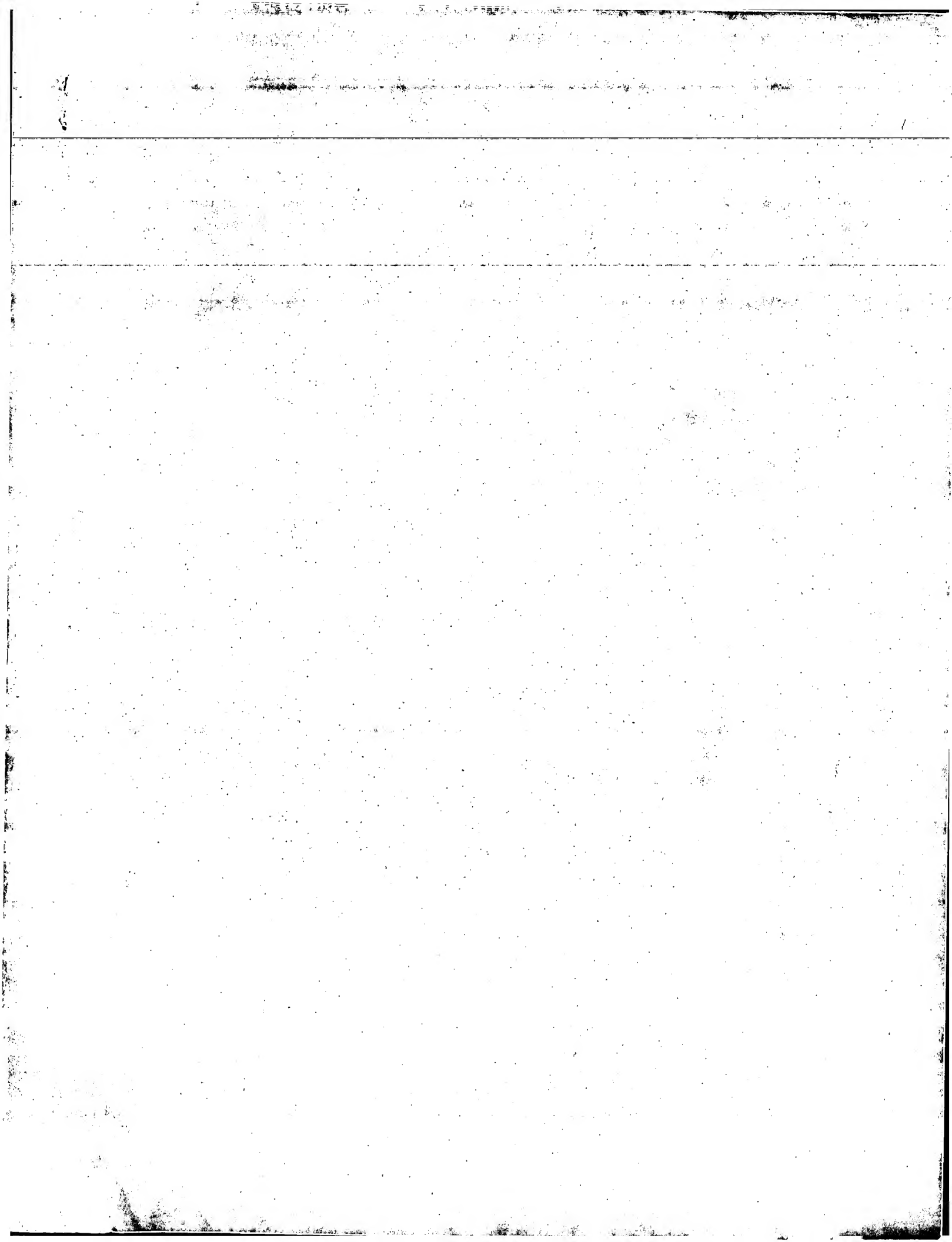
RESULT 9
US-10-156-761-1/C
Sequence 1, Application US/10156761
Publication No. US20030119018A1

GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIRAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1

SEX ID NO 4025

SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 4625



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:13:45 ; Search time 503.434 Seconds

(without alignments)
900.761 Million cell updates/sec

Title: US-10-087-631b-9

Perfect score: 28

Sequence: 1 cgtcgtggagatagcgcgcacgtgtt 28

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	20	71.4	475	13	BI611725
c 2	19	67.9	406	10	AV732708
c 3	19	67.9	590	17	A2567844
c 4	19	67.9	649	13	BMS97959
c 5	19	67.9	667	9	AI055301
c 6	19	67.9	900	12	BF620238

c 7	18.6	66.4	521	17	A0350115
c 8	18.6	66.4	549	17	BH282580
c 9	18.6	66.4	552	17	A0776894
c 10	18.6	66.4	560	17	A2397654
c 11	18.6	66.4	635	17	BH265153
c 12	18.6	66.4	640	17	A2291296
c 13	18.6	66.4	736	12	BG293648
c 14	18.6	66.4	763	17	BH267142
c 15	18.6	66.4	805	17	BH279359
c 16	18.6	66.4	865	17	A0740087
c 17	18.6	66.4	868	13	BI761876
c 18	18.4	65.7	131	10	BE148040
c 19	18.4	65.7	247	12	BG410222
c 20	18.4	65.7	265	12	BG410420
c 21	18.4	65.7	289	10	BH348944
c 22	18.4	65.7	325	10	AM566217
c 23	18.4	65.7	400	10	AM528213
c 24	18.4	65.7	480	9	AA820730
c 25	18.4	65.7	482	14	BQ996871
c 26	18.4	65.7	513	12	BE978187
c 27	18.4	65.7	529	17	AQ913537
c 28	18.4	65.7	533	13	BI167853
c 29	18.4	65.7	541	13	BI163484
c 30	18.4	65.7	560	13	BI370216
c 31	18.4	65.7	573	13	BI229543
c 32	18.4	65.7	593	13	BI632278
c 33	18.4	65.7	605	9	AT520529
c 34	18.4	65.7	613	9	AA568072
c 35	18.4	65.7	620	14	BQ988356
c 36	18.4	65.7	652	14	BQ999054
c 37	18.4	65.7	662	13	BI636271
c 38	18.4	65.7	752	12	A2132792
c 39	18.4	65.7	769	17	A2132929
c 40	18.4	65.7	812	12	BG435911
c 41	18.4	65.7	1064	14	BQ048851
c 42	18.2	65.0	757	17	AG185522
c 43	18.2	65.0	846	12	BF138165
c 44	18.2	65.0	1033	12	BG390990
c 45	18.2	64.3	166	9	AV058751

ALIGNMENTS

RESULT 1
LOCUS BI611725/c 475 bp mRNA linear EST 07-SEP-2001
DEFINITION RH18118.5prtime RH Drosophila melanogaster normalized Head pfc1-1
Drosophila melanogaster cDNA clone RH18118 5 similar to sqd:
F8an0017791 GO:RNA binding (GO:0003723); RNA binding (GO:0003723)]
located on: 3R 87F7-87F7; 08/17/2001, mRNA sequence.

ACCESSION BI611725.1 GI:15507250
VERSION BI611725.1
KEYWORDS fruit fly
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 475)
J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Mista,S.,
Mungall,C.J., Nunoo,J., Paclib,J., Paragas,V., Park,S.,
Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celisner,S. and Rubin
G.M.

TITLE BDGP/HMI RH Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, esl@fruitfly.berkeley.edu
hit genomic AEO03701: arm:3R (935746,956163)
estimated-cyt:87F4-87F4: 08/17/2001
Plate: RH 181 row: B Column: 6
High quality sequence stop: 291.

FEATURES

Location/Qualifiers

1..475

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="RH1818"

/clone_lib="RH Drosophila melanogaster normalized Head

/sex="male and female"

/dev_stage="Adult"

/lab_host="DHS-alpha Tona"

/note="Organ: head; Vector: pELC1; Site:1: XhoI; Site:2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized by Piero Carninci at Cre recombinase. Plasmid CDNA library."

150 a 108 c 110 g 107 t

BASE COUNT

150 a 108 c 110 g 107 t

Query Match

Best Local Similarity 71.4%; Score 20; DB 13; Length 475;

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY

1 CGTCTGGGATAGTCGTCATGCTT 28

347 CCTCTGGGATAGTCGTCATGCTT 320

Db

RESULT 2
AV732708

LOCUS AV732708 406 bp mRNA linear EST 17-OCT-2000

DEFINITION AV732708 HTF Homo sapiens CDNA clone HTFCH01 5', mRNA sequence.

ACCESSION AV732708

VERSION AV732708.1

KEYWORDS GI:10850253

SOURCE EST.

ORGANISM human.

REFERENCE Homo sapiens

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 406)

Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,

Li, N., Qian, B., Liu, J., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu

, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,

Chen, J., Chen, Z., and Han, Z.

Homo sapiens CDNA HTF clones

Unpublished (2000)

COMMENT Contact: Zenguan Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919(ex.45)

Fax: 86-21-50801922

Email: hanzg@hgc.sh.cn

This clone is available at CHGC in Shanghai.

Location/Qualifiers

1..406

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="HTFCH01"

/clone_lib="HTF"

/tissue_type="Hypothalamus"

/dev_stage="Adult"

/lab_host="SOLR"

/note="Vector: pBluescript SK(-); Site:1: EcoRI; Site:2: XhoI"

BASE COUNT

153 a 50 c 65 g 138 t

ORIGIN

Query Match

Best Local Similarity 67.9%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Query Match

Best Local Similarity 81.5%; Score 19; DB 10; Length 406;

Matches 81.5%; Pred. No. 3.9e+02;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 CGTCTGGGATAGTCGTCATGCTT 27

376 CGATGCGGTATAGTACGTCATGCTT 402

Db

RESULT 3

LOCUS A2567844

DEFINITION 239PvH07 Pv MEN #30 Plasmodium vivax genomic 3', DNA sequence.

ACCESSION A2567844

VERSION A2567844.1

KEYWORDS GI:13976352

SOURCE GSS.

ORGANISM malaria parasite P. vivax.

Plasmodium vivax

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

1 (bases 1 to 590)

Carlton, J.M., R. and Dame, J.B.

The Plasmodium vivax and P. berghei gene sequence tag projects

Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)

Contact: Dame JB

Dept. of Pathobiology, College of Veterinary Medicine

University of Florida

2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA

Tel: 352 392 4700

Fax: 352 392 9704

Email: damej@mail.vetmed.ufl.edu

Seq primer: M13(-20) forward

Class: Shotgun.

Location/Qualifiers

1..590

/organism="Plasmodium vivax"

/strain="Salvador I (Collins, W. 1972. J. Parasitol. 69,

497-598)."

/db_xref="taxon:5855"

/clone_lib="Pv MEN #30"

/dev_stage="asexual blood forms"

/lab_host="Salintri bolivensis"

/note="Vector: pBluescript SK(+)"

Host leukocytes were extracted from P. vivax infected

blood using the following methods: first, infected blood

per 10 ml blood. Then blood was passed over a column of

acid washed 0.1 mm glass beads, then through a Plasmidipur

filter, followed by passage through a column of pre-wet

Whatman CF11 powder (1:2 ratio volume of blood to CF11),

and finally centrifuged through a 50% Percoll density

cushion. Purified DNA was digested with mung bean nuclease

(Verneick, K.D., Imberski, R.B., and McCutchan, T.F. 1988.

Nucleic Acids Research 16:6883-6896). Digested DNA was

blunt-ended using T4 DNA Polymerase and size fractionated

over a Sepharose CL-2B column. Fractions in the size range

500bp-4kb were ligated into the Eco RV site of pBluescript

SK(+), and E. coli XL-10 Gold transformed with the

ligation mixture."

BASE COUNT

142 a 138 c 155 g 155 t

ORIGIN

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match

Best Local Similarity 67.9%; Score 19; DB 17; Length 590;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

</

DEFINITION 17000687506516 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
 ACCESSION 19600449733180 5', mRNA sequence.
 VERSION BM597959
 KEYWORDS EST.
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
 REFERENCE 1 (bases 1 to 649)
 AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
 TITLE R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
 JOURNAL Celera Anopheles gambiae EST project
 COMMENT Unpublished (2002)
 CONTACT: Holt R.A.
 Celera Genomics
 45 W. Gude Dr., Rockville, MD 20850, USA
 Tel: 2404534580
 Fax: 2404534580
 Email: Holt@celera.com
 Plate: NU010049UT row: O column: 22
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..649
 /organism="Anopheles gambiae"
 /strain="RSP-ST (Reduced susc. to Permethrin - std.
 /chromosome)
 /db_xref="taxon:7165"
 /clone="19600449733180"
 /clone_1ib="A.Gam.ad.cDNA.blood1"
 /dev_stage="Adult"
 /lab_host="DH10b"
 /note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen 24 hours after human blood feeding. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)"
 BASE COUNT 143 a 185 c 186 g 135 t
 ORIGIN
 Query Match 67.9%; Score 19; DB 13; Length 649;
 Best Local Similarity 81.5%; Pred. No. 4.7e+02;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 CGTTCGTGGATAGTCCTCATGCTGT 27
 11 ||||| ||||| ||||| ||||| |||||
 11 CGTTCGTGGATAGTCCTCATGCTGT 37
 DB 11 CGTTCGTGGATAGTCCTCATGCTGT 37
 RESULT 5
 LOCUS A1055301 667 bp mRNA linear EST 16-JUL-1998
 DEFINITION coau0003j18 Cotton Boll Abscission zone cDNA library Gossypium
 ACCESSION A1055301
 VERSION A1055301
 KEYWORDS EST.
 SOURCE upland cotton.
 ORGANISM Gossypium hirsutum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
 REFERENCE 1 (bases 1 to 667)
 AUTHORS Wan,C.-H., Yu,Y., Sasnowski,M. and Wang,R.A.
 TITLE Cotton EST Database: Sequence Analysis of 2000 cDNA clones from an
 JOURNAL Abscission zone library
 COMMENT Unpublished (1998)
 CONTACT: Wang RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA

FEATURES
 source
 1..667
 /organism="Gossypium hirsutum"
 /db_xref="taxon:3635"
 /clone="coau0003j18"
 /clone_1ib="Cotton Boll Abscission zone cDNA library"
 /lab_host="XLI-Blue MRF"
 /note="Vector: Uni-ZAP XR; Site 1: EcoRI; Site 2: XhoI; This is a Uni-ZAP XR custom cDNA library made by Stratagene (U.S.A.: 1-800-424-5444); Stratagene cat. #837201."
 BASE COUNT 176 a 141 c 149 g 199 t 2 others
 ORIGIN
 Query Match 67.9%; Score 19; DB 9; Length 667;
 Best Local Similarity 81.5%; Pred. No. 4.7e+02;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 CGTTCGTGGATAGTCCTCATGCTGT 27
 11 ||||| ||||| ||||| ||||| |||||
 45 CGTTCGTGGATAGTCCTCATGCTGT 71
 DB 45 CGTTCGTGGATAGTCCTCATGCTGT 71
 RESULT 6
 LOCUS BF620238/c 900 bp mRNA linear EST 22-OCT-2001
 DEFINITION HVSMEC0019C13f Hordeum vulgare seedling shoot EST library
 HVCNMA0003 (Etiolated and unstressed) Hordeum vulgare cDNA clone
 HVSMEC0019C13f, mRNA sequence.
 ACCESSION BF620238
 VERSION BF620238
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 /trillceae; Hordeum.
 1 (bases 1 to 900)
 Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
 Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
 R.D., Oates,R. and Main,D.
 Development of a genetically and physically anchored EST resource
 for barley genomics: Morex unstressed seedling shoot cDNA library
 Unpublished (2001)
 On Dec 18, 2000 this sequence version replaced gi:11883972.
 CONTACT: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Total hg bases - 361
 Seq primer: AATTACCTCACTAAGG
 High quality sequence stop: 671.
 Location/Qualifiers
 1..900
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSMEC0019C13f"
 /clone_1ib="Hordeum vulgare seedling shoot EST library
 HVCNMA0003 (Etiolated and unstressed)"
 /tissue_type="Seedling shoot"
 /lab_host="TDC121"
 /note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
 Seeds were surface sterilized then germinated under axenic
 conditions in the dark at room temperature on filter paper

with water, myristatin and cefotaxime in covered crystallization dishes. Five-day old seedling shoots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million plv were in vivo excised to give bluescript SK(-) cDNA phagemids. These steps were performed in the Tu riverside laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close, T, Wing, R, Klein, A, Wise, R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/g9pages/bgn/31/cover.html>)

BASE COUNT
ORIGIN
Query Match
Best Local Similarity 67.9% Score 19; DB 12; Length 900;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db 1 CGTTCGTGGATGATCGTCATGCTGT 27
464 CGTTCGTGGATGATCGTCATGCTGT 438

RESULT 7
LOCUS AG350115
DEFINITION RPII11-110C18.TV RPII-11 Homo sapiens genomic clone RPII-11-110C18,
DNA sequence.
ACCESSION AG350115
VERSION AG350115.1 GI:4177450
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Zhuo, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter
J.C.C.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Use of BAC End Sequences from Library RPII-11 for Sequence-Ready
Map Building
Unpublished (1997)
Other GSSs: RPII11-110C18.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetlgr.org

Clones are derived from the human BAC library RPII-11. For BAC
library availability, please contact Pieter de Jong
BACPAC Resources (<http://bacpac.med.buffalo.edu>). Clones may be purchased from
Research Genetics (<http://inforesgen.com>). BAC end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.

FEATURES
SOURCE
Location/Qualifiers
1..521

/organism="Homo sapiens"
/db_xref="GDB:7541921"
/db_xref="taxon:9606"

/clone="RPII-11-110C18"
/clone_11b="RPII-11"
/sex="Male"
/cell_type="lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPII11 Human Male BAC library"

BASE COUNT
ORIGIN
Query Match
Best Local Similarity 66.4% Score 18.6; DB 17; Length 521;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Db 3 TTCTGGGATGATCGTCATGCTGT 27
254 TTCTGGGATGATGATGCTGT 230

RESULT 8
LOCUS BH282580/c
DEFINITION CH230-12818.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
ACCESSION BH282580
VERSION BH282580.1 GI:17194982
KEYWORDS GSS.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 549)
Zhuo, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shartsbeyn
, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de
Jong, P. and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 EcoRI segment
Unpublished (1999)
Other GSSs: CH230-12818.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library
availability, please contact Pieter de Jong (pdejong@tigr.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering_information.htm). BAC end
plate: 128 row: 1 column: 8
Seq primer: SP6
Class: BAC ends.

FEATURES
SOURCE
Location/Qualifiers
1..549

/organism="Rattus norvegicus"
/strain="BN/SSNsd/MCw"
/db_xref="taxon:10116"
/clone="CH230-12818"
/clone_11b="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
CHORI-230 Rat (BN/SSNsd/MCw) BAC library produced by
Pieter de Jong"

BASE COUNT
ORIGIN
Query Match
Best Local Similarity 66.4% Score 18.6; DB 17; Length 549;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GTTCGTGGATAGTCCGTCATGCTG 26
 DB 101 GTTAGAGGATAGTCACTAGTATGCTG 77

RESULT 9
 LOCUS AQ776894/c 552 bp DNA linear GSS 02-AUG-1999
 DEFINITION HS_5310.A2.B07.T7A.RPCT-11 Human Male BAC Library Homo sapiens
 genomic clone Plate=886 Col=14 Row=C, DNA sequence:
 ACCESSION AQ776894
 VERSION AQ776894.1 GI:5679854
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 552)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,R., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCT-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPec Resources (http://bacpec.med.buffalo.edu/ordering_bac.htm)
 or from Resear h Genetics (info@resgen.com). BAC end Web Server:
 http://www.htsc.washington.edu
 Plate: 886 row: C column: 14
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 552.

FEATURES
 SOURCE
 Location/Qualifiers
 1..552
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_plate="886 Col=14 Row=C"
 /clone_lib="RPCT-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRII. Size selected DNA was cloned into the
 pBACe3.6 vector at EcoRI sites"

BASE COUNT 201 a 114 c 92 g 139 t 6 others

ORIGIN
 Query Match 66.4%; Score 18.6; DB 17; Length 552;
 Best Local Similarity 84.0%; Pred. NO. 6.6e02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TTTCGTGGATAGTCCGTCATGCTG 27
 DB 286 TTTCGTGGATAGTCCGTCATGCTG 262

RESULT 10
 LOCUS A2397654 560 bp DNA linear GSS 03-OCT-2000
 DEFINITION IM0162D13R Mouse 10kb plasmid U0GC1M library Mus musculus genomic
 clone U0GC1M0162D13 R, DNA sequence.
 ACCESSION A2397654
 VERSION A2397654.1 GI:10512726.

KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 560)
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss,R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0162 row: D column: 13
 Seq primer: CACACAGCAACACCTAGACC
 Class: plasmid ends
 High quality sequence stop: 560.

FEATURES
 SOURCE
 Location/Qualifiers
 1..560
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U0GC1M0162D13"
 /clone_lib="Mouse 10kb plasmid U0GC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42mv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repated with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g14732114[gb]AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 118 a 106 c 156 g 180 t

ORIGIN
 Query Match 66.4%; Score 18.6; DB 17; Length 560;
 Best Local Similarity 84.0%; Pred. NO. 6.6e02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TTTCGTGGATAGTCCGTCATGCTG 27
 DB 2 TTTCGTGGATAGTCCGTCATGCTG 26

RESULT 11
 LOCUS BH265153 635 bp DNA linear GSS 30-NOV-2001
 DEFINITION CH230-55121.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
 BH265153
 ACCESSION BH265153
 VERSION BH265153.1 GI:17174769

KEYWORDS
SOURCEGSS.
Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus

1 (bases 1 to 635)

Zhaio, S., Shetty, J., Shatsman, S., Tsengaye, G., Geer, K., Shvartsbeyn

A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Nigys, F., de

Jong, P. and Fraser, C.M.

Rat BAC End Sequences from Library CHORI-230 EcorI segment

Unpublished (1999)

Other GSS: CH230-55121.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhaio@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tftp/bac_ends/mouse/bac_end_intro.html

Plate: 131 row: N column: 12

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. 640

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-131N12"

/clone_1id="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1;

EcorI; Site: 2: EcorI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcorI and EcorI Methylase. Size

selected DNA was cloned into the pBACe3.6 vector at the

EcorI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies).

BASE COUNT

163 a 142 c 157 g 178 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

231 a 117 c 81 g 206 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

231 a 117 c 81 g 206 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

231 a 117 c 81 g 206 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

231 a 117 c 81 g 206 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

231 a 117 c 81 g 206 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

231 a 117 c 81 g 206 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

231 a 117 c 81 g 206 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

231 a 117 c 81 g 206 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

231 a 117 c 81 g 206 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

231 a 117 c 81 g 206 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

231 a 117 c 81 g 206 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

231 a 117 c 81 g 206 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

231 a 117 c 81 g 206 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

231 a 117 c 81 g 206 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

231 a 117 c 81 g 206 t

ORIGIN

1. 635

/organism="Rattus norvegicus"

/strain="BN/SSNhsd/MCv"

/db_xref="taxon:10116"

/clone="CH230-55121"

/clone_1id="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site: 1: EcorI; Site: 2: EcorI;

CHORI-230 Rat (BN/SSNhsd/MCv) BAC library produced by

Pieter de Jong"

Plater de Jong"

BASE COUNT

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_image="4502619"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally; oligo-dt primed. Average insert size 3.3 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT      181 a      160 c      210 g      185 t
ORIGIN

Query Match      66.4%; Score 18.6; DB 12; Length 736;
Best Local Similarity 84.0%; Pred. No. 7.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      3 TTGCTGGATGATCGCTCATGCTGT 27
        ||||| ||||| ||||| |||||
        2 TTGCTGGATGATCGCTCATGCTGT 26

RESULT 14
BH267142/c      763 bp      DNA      linear      GSS 30-NOV-2001
LOCUS           CH230-55010.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
DEFINITION      BH267142
ACCESSION       BH267142
VERSION         BH267142.1 GI:11719062
KEYWORDS        GSS.
SOURCE          Norway rat.
ORGANISM        Rattus norvegicus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                Rattus.
                1 (bases 1 to 763).
                Zhao, S., Shetty, J., Shatsman, S., Teegaye, G., Geer, K., Shvartsbeyn
                A., Gebregorjiss, E., Overton, L., Russell, D., Chen, D., Riggs, F., de
                Jong, P. and Fraser, C.M.
                Rat BAC End Sequences from Library CHORI-230 EcORI segment
                Unpublished (1999)
                Other GSSs: CH230-55010.TV
                Contact: Shaying Zhao
                Department of Eukaryotic Genomics
                The Institute for Genomic Research
                9712 Medical Center Dr., Rockville, MD 20850, USA
                Tel: 301 838 0200
                Fax: 301 838 0208
                Email: szhao@tigr.org
                Clones are derived from the rat BAC library CHORI-230
                (http://www.chori.org/bacpac/rat230.htm). For BAC library
                availability, please contact Pieter de Jong (pdejong@email.cho.org).
                Clones may be purchased from BACPAC Resources
                (http://www.chori.org/bacpac/or_ering_information.htm). BAC end
                page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
                Plate: 55 row: 0 column: 10
                Seq primer: SP6
                Class: BAC ends.

FEATURES
Source
Location/Qualifiers
1..763
/organism="Rattus norvegicus"
/strain="BN/SSNhd/MCW"
/db_xref="taxon:10116"
/clone="CH230-55010"
/clone_lib="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/notes="Vector: pTARAC2.1; Site: 1: EcoRI; Site: 2: EcoRI;
CHORI-230 Rat (BN/SSNhd/MCW) BAC library produced by
Pieter de Jong"

BASE COUNT      275 a      139 c      98 g      251 t
ORIGIN

```

```

Query Match      66.4%; Score 18.6; DB 17; Length 805;
Best Local Similarity 84.0%; Pred. No. 7.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      2 GTTCGTGGATGATCGCTCATGCTG 26
        ||| | ||||| || |||||
        104 GTTAGAGGATGATGATGATGCTG 80

RESULT 15
BH279359/c      805 bp      DNA      linear      GSS 30-NOV-2001
LOCUS           CH230-128C8.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
DEFINITION      BH279359
ACCESSION       BH279359
VERSION         BH279359.1 GI:117191761
KEYWORDS        GSS.
SOURCE          Norway rat.
ORGANISM        Rattus norvegicus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                Rattus.
                1 (bases 1 to 805)
                Zhao, S., Shetty, J., Shatsman, S., Teegaye, G., Geer, K., Shvartsbeyn
                A., Gebregorjiss, E., Overton, L., Russell, D., Chen, D., Riggs, F., de
                Jong, P. and Fraser, C.M.
                Rat BAC End Sequences from Library CHORI-230 EcORI segment
                Unpublished (1999)
                Contact: Shaying Zhao
                Department of Eukaryotic Genomics
                The Institute for Genomic Research
                9712 Medical Center Dr., Rockville, MD 20850, USA
                Tel: 301 838 0200
                Fax: 301 838 0208
                Email: szhao@tigr.org
                Clones are derived from the rat BAC library CHORI-230
                (http://www.chori.org/bacpac/rat230.htm). For BAC library
                availability, please contact Pieter de Jong (pdejong@email.cho.org).
                Clones may be purchased from BACPAC Resources
                (http://www.chori.org/bacpac/or_ering_information.htm). BAC end
                page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
                Plate: 128 row: C column: 8
                Seq primer: SP6
                Class: BAC ends.

FEATURES
Source
Location/Qualifiers
1..805
/organism="Rattus norvegicus"
/strain="BN/SSNhd/MCW"
/db_xref="taxon:10116"
/clone="CH230-128C8"
/clone_lib="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/notes="Vector: pTARAC2.1; Site: 1: EcoRI; Site: 2: EcoRI;
CHORI-230 Rat (BN/SSNhd/MCW) BAC library produced by
Pieter de Jong"

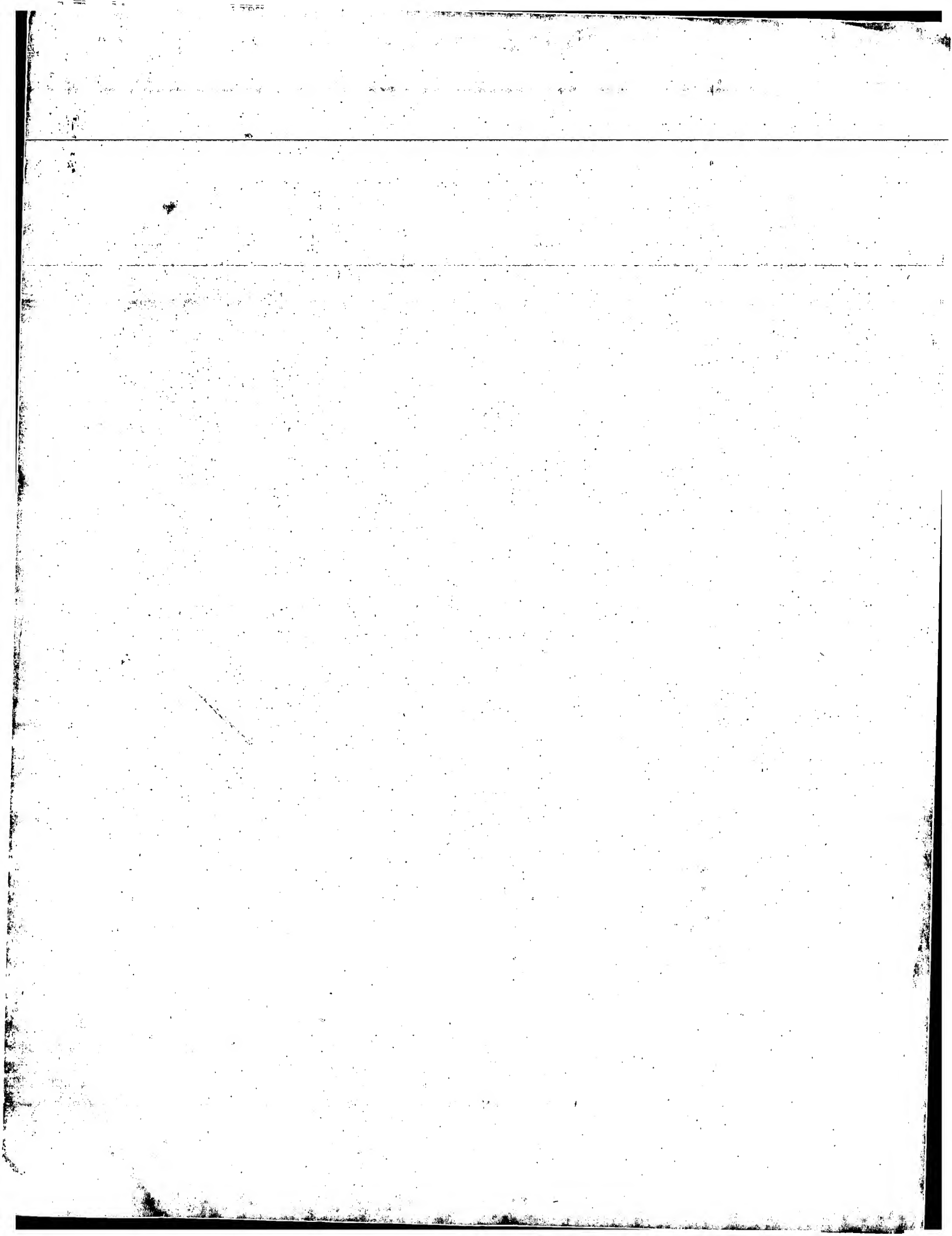
BASE COUNT      284 a      139 c      109 g      273 t
ORIGIN

Query Match      66.4%; Score 18.6; DB 17; Length 805;
Best Local Similarity 84.0%; Pred. No. 7.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      2 GTTCGTGGATGATCGCTCATGCTG 26
        ||| | ||||| || |||||
        113 GTTAGAGGATGATGATGATGCTG 89

Search completed: July 11, 2003, 02:26:12
Job time : 510.434 secs

```



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 259.955 Seconds

(without alignments)
3358.597 Million cell updates/sec

Title: US-10-087-631B-4

Perfect score: 30

Sequence: 1 tggactcagccttgctgcctcctctt 30

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hlg.*
3: gb_in.*
4: gb_cm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sy.*
12: gb_un.*
13: gb_vl.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_hlg_hum.*
31: em_hlg_inv.*
32: em_hlg_other.*
33: em_hlg_mus.*
34: em_hlg_pln.*
35: em_hlg_rod.*
36: em_hlg_mam.*
37: em_hlg_vrt.*
38: em_sy.*
39: em_hlgo_hum.*
40: em_hlgo_mus.*
41: em_hlgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21.2	70.7	161776	AL133229	AL133229 Human DNA
2	21.2	70.7	166523	AC022646	AC022646 Homo sapi
3	21.2	70.7	198917	AL356157	AL356157 Human DNA
4	21	70.0	6751	AF408636	AF408636 Bacterioph
5	21	70.0	30265	AC013516	AC013516 Homo sapi
6	21	70.0	93312	HS32110	HS32110 Human DNA s
7	21	70.0	141313	AC004943	AC004943 Homo sapi
8	21	70.0	147091	AC108030	AC108030 Homo sapi
9	21	70.0	151552	AC024132	AC024132 Homo sapi
10	21	70.0	156128	AC121186	AC121186 Rattus no
11	21	70.0	160696	AC090937	AC090937 Homo sapi
12	21	70.0	161604	AP001978	AP001978 Homo sapi
13	21	70.0	177876	AC025163	AC025163 Homo sapi
14	21	70.0	182695	AC130278	AC130278 Mus muscu
15	21	70.0	193700	AC024545	AC024545 Homo sapi
16	21	70.0	227461	AC021429	AC021429 Homo sapi
17	20.6	68.7	117890	AL133466	AL133466 Human DNA
18	20.6	68.7	148851	AC016227	AC016227 Homo sapi
19	20.6	68.7	152926	AC098093	AC098093 Rattus no
20	20.6	68.7	160599	AC096630	AC096630 Pan trogl
21	20.6	68.7	164435	AC112508	AC112508 Homo sapi
22	20.6	68.7	178089	AC010929	AC010929 Homo sapi
23	20.6	68.7	179604	AP002959	AP002959 Homo sapi
24	20.6	68.7	185319	AL844848	AL844848 Mus muscu
25	20.6	68.7	186111	AC096341	AC096341 Rattus no
26	20.6	68.7	187917	AC025321	AC025321 Homo sapi
27	20.6	68.7	193677	AC112860	AC112860 Rattus no
28	20.6	68.7	209021	AC098878	AC098878 Mus muscu
29	20.4	68.0	2077	AK057308	AK057308 Homo sapi
30	20.4	68.0	152422	AC016491	AC016491 Homo sapi
31	20.4	68.0	154433	AL732447	AL732447 Mouse DNA
32	20.4	68.0	154936	AC021252	AC021252 Homo sapi
33	20.4	68.0	156848	AC097710	AC097710 Homo sapi
34	20.4	68.0	158079	AC122129	AC122129 Homo sapi
35	20.4	68.0	160544	HS25307	AL354000 Homo sapi
36	20.4	68.0	160918	AC069196	AC069196 Homo sapi
37	20.4	68.0	162584	AC018594	AC018594 Homo sapi
38	20.4	68.0	164519	AC020630	AC020630 Homo sapi
39	20.4	68.0	168720	AC092987	AC092987 Homo sapi
40	20.4	68.0	170814	AL844565	AL844565 Mus muscu
41	20.4	68.0	171942	AC021715	AC021715 Homo sapi
42	20.4	68.0	175948	AC019192	AC019192 Homo sapi
43	20.4	68.0	176688	AC084696	AC084696 Mus muscu
44	20.4	68.0	181899	AC015904	AC015904 Homo sapi
45	20.4	68.0	181905	CNS01R1B	AL163052 Human chr

ALIGNMENTS

RESULT 1
LOCUS AL133229
DEFINITION Human DNA sequence from clone RPS-1167E19 on chromosome 20 Contains
STES, GSSs and a Cpg island, complete sequence.
ACCESSION AL133229
VERSION AL133229.40 GI:8546579
KEYWORDS HTG; Cpg Island.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 161776)
Bates, K.
Direct Submission

JOURNAL

COMMENT

Submitted (13-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 On June 14, 2000 this sequence version replaced g1:8517286.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
<http://www.sanger.ac.uk/projects/C.elegans/wormpep>
 This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr20>
 This sequence is the entire insert of clone RP5-1167E19. The true left end of clone RP11-179J15 is at 93503 in this sequence. The true right end of clone RP3-50802 is at 149816 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-1167E19 is from the library RPI-5 constructed by the group of Pieter de Jong. For further details see
<http://www.ncbi.nlm.nih.gov/bacpac/home.htm>
 VECTOR: pCYPAC2.

FEATURES

SOURCE

Location/Qualifiers
 1..161776
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /clone="RP5-1167E19"
 /clone_1bp="RPI-5"
 145..449
 /note="match: STS: Em:267317"
 complement(172..577)
 repeat_region
 278..333
 /note="match: GSS: Em:AQ729312"
 repeat_region
 720..756
 /note="28 copies 2 mer ag 76% conserved"
 repeat_region
 1741..1785
 /note="12 repeat: matches 2379..2415 of consensus"
 repeat_region
 1807..1893
 /note="15 copies 3 mer ata 100% conserved"
 repeat_region
 2601..2693
 /note="MIR repeat: matches 39..122 of consensus"
 repeat_region
 2743..2880
 /note="HAL1 repeat: matches 5..98 of consensus"
 repeat_region
 3243..3403
 /note="LTR45 repeat: matches 389..525 of consensus"
 repeat_region
 3404..3703
 /note="HAL1 repeat: matches 425..559 of consensus"
 repeat_region
 complement(3631..4052)
 /note="Alusq repeat: matches 1..300 of consensus"
 misc_feature
 /note="match: GSS: Em:B95838"
 complement(3631..4052)
 /note="match: GSS: Em:AQ010447"
 3704..3758
 /note="HAL1 repeat: matches 559..618 of consensus"
 repeat_region
 3956..4352
 /note="HAL1 repeat: matches 360..749 of consensus"
 repeat_region
 4353..4643
 /note="HAL1 repeat: matches 360..749 of consensus"
 repeat_region
 4644..4708
 /note="Alusx repeat: matches 1..290 of consensus"
 repeat_region
 6054..6145
 /note="HAL1 repeat: matches 295..360 of consensus"

repeat_region
 6528..6652
 /note="MIR repeat: matches 93..197 of consensus"
 repeat_region
 7379..7531
 /note="MIR repeat: matches 47..169 of consensus"
 repeat_region
 8058..8267
 /note="MIR repeat: matches 27..188 of consensus"
 repeat_region
 8631..8827
 /note="MIR repeat: matches 12..262 of consensus"
 repeat_region
 8828..9148
 /note="12 repeat: matches 739..939 of consensus"
 repeat_region
 9149..9501
 /note="Alu repeat: matches 1..309 of consensus"
 repeat_region
 9507..9559
 /note="L2 repeat: matches 939..1277 of consensus"
 repeat_region
 9595..9883
 /note="Alub repeat: matches 249..301 of consensus"
 repeat_region
 9884..10018
 /note="Alub repeat: matches 1..252 of consensus"
 repeat_region
 10076..10172
 /note="L2 repeat: matches 1262..1398 of consensus"
 repeat_region
 10446..10663
 /note="MIR1J repeat: matches 104..200 of consensus"
 repeat_region
 10824..10911
 /note="LTR33 repeat: matches 294..521 of consensus"
 repeat_region
 10912..11493
 /note="MIR1I repeat: matches 317..410 of consensus"
 repeat_region
 11505..12032
 /note="L2 repeat: matches 1396..2019 of consensus"
 repeat_region
 12058..12368
 /note="L2 repeat: matches 2208..2738 of consensus"
 repeat_region
 complement(12727..13254)
 /note="Alusq repeat: matches 1..311 of consensus"
 misc_feature
 /note="match: GSS: Em:AQ714903"
 13528..13670
 /note="L2 repeat: matches 1293..1445 of consensus"
 repeat_region
 13697..13763
 /note="L2 repeat: matches 2607..2674 of consensus"
 repeat_region
 14119..14269
 /note="L2 repeat: matches 2607..2674 of consensus"
 repeat_region
 14432..14650
 /note="LIM4 repeat: matches 4384..4543 of consensus"
 repeat_region
 14802..15103
 /note="LIMC/D repeat: matches 5428..5642 of consensus"
 repeat_region
 15307..15725
 /note="Alus repeat: matches 2..303 of consensus"
 repeat_region
 15745..16662
 /note="MIR1J repeat: matches 68..507 of consensus"
 repeat_region
 16794..16875
 /note="LIM4 repeat: matches 3730..4663 of consensus"
 repeat_region
 16876..17183
 /note="LIM4 repeat: matches 3460..3538 of consensus"
 repeat_region
 17184..17510
 /note="Alu repeat: matches 1..304 of consensus"
 repeat_region
 17439..17841
 /note="LIM4 repeat: matches 3115..3460 of consensus"
 misc_feature
 /note="match: GSS: Em:B32565"
 17511..17611
 /note="HAL1 repeat: matches 707..813 of consensus"
 repeat_region
 complement(17614..18091)
 /note="match: GSS: Em:AQ264589"
 17689..17892
 /note="LIM3 repeat: matches 5524..5728 of consensus"
 repeat_region
 18401..18707
 /note="Alusq repeat: matches 1..308 of consensus"
 repeat_region
 19010..19193
 /note="MIR1J repeat: matches 3..232 of consensus"
 repeat_region
 20676..20803
 /note="64 copies 2 mer ag 62% conserved"
 misc_feature
 21194..21437
 /note="match: GSS: Em:AQ834343"
 22509..22811
 /note="Alu repeat: matches 1..297 of consensus"
 repeat_region
 complement(23846..24267)
 /note="match: GSS: Em:AQ562784"

[illegible]

Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
 Gardany, S., Grant, G., Hagos, B., Heathford, A., Horton, L.,
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karats, A., Klein, J.,
 Landers, T. P., Leback, J., Levine, R., Liew, C., Liu, G., Locke, K.,
 Macdonald, P., Marquis, N., McKean, P., McGirk, A., McInerlan, K.,
 McPheters, R., Medlert, T., Menus, L., Morrow, J., Naylor, J.,
 Norman, C. H., O'Connor, T., O'Donnell, P., Olivar, T. M., Peterson, K.,
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,
 Tiell, A., Vassiliou, H., Viel, R., Vo, A., Wu, X., Wymann, D., Ye, W., J.
 Zimmerman, A. and Zody, M.
 Direct Submission
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 16, 2000 this sequence version replaced gi:5984422.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 Genome Center

```

Center Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www.seg.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----
Project Information
Center project name: I4741
Center clone name: 28_C.18
-----
Summary Statistics
Sequencing Summary: M13: M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 157134 bases at least Q40
Consensus quality: 161890 bases at least Q30
Consensus quality: 163469 bases at least Q20
Insert size: 156000; agarose-fp
Insert size: 164823; sum-of-contigs
Quality coverage: 5.2 in Q20 bases; agarose-fp
Quality coverage: 4.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1100 1199: gap of 1099 bp in length
1200 2705: contig of 1506 bp in length
2706 2805: gap of 100 bp
2806 4225: contig of 1420 bp in length
4226 4325: gap of 100 bp
4326 5737: contig of 1412 bp in length
5738 5837: gap of 100 bp
5838 7280: contig of 1443 bp in length
7281 7380: gap of 100 bp
7381 11595: contig of 4215 bp in length
11596 11695: gap of 100 bp
11696 16125: contig of 4430 bp in length
16126 16225: gap of 100 bp
16226 21190: contig of 4965 bp in length
21191 21290: gap of 100 bp
21291 27715: contig of 6425 bp in length
27716 27815: gap of 100 bp
27816 33432: contig of 5617 bp in length
33433 33532: gap of 100 bp
33533 46593: contig of 13061 bp in length
46594 46693: gap of 100 bp
46694 58185: contig of 11492 bp in length
58186 58285: gap of 100 bp
58286 70164: contig of 11679 bp in length
70165 70264: gap of 100 bp

```

FEATURES
source

COMMENT

```

FEATURES
    source
        location/Qualifiers
            1..198917
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="10"
                /clone="Rp11-733D4"
                /clone_idb="RPCT1-11.3"

```

```

repeat_region      /note="U2 repeat: matches 2. .36 of consensus"
                    1059. .1224
                    /note="MIR repeat: matches 33. .195 of consensus"
misc_features      2002

```

```
repeat_region      /note="match: GSS: Em:AQ103508"  
                    3623. .3937  
repeat_region      /note="AluX repeat: matches 1. .312 of consensus"  
                    4292. .4540
```

```

repeat_region      8422. 10552      2621. 4014 of consensus"
repeat_region      10553. 11064      4101. 6155 of consensus"

```

```

repeat_region 11065..11767
                /note="LIP1 repeat: matches 3424..4101 of consensus"
repeat_region 11768..12024
                /note="AUG5 repeat: matches 54..310 of consensus"
repeat_region 12025..14589
                /note="LIP1 repeat: matches -243..3424 of consensus"
                /note="LIP1 repeat: matches -243..3424 of consensus"

```

```

repeat_region 14608..14681
/note="LIPB1 repeat: matches -485..-412 of consensus"
repeat_region 14732..14930
/note="LIPBa repeat: matches -1543..-1345 of consensus"
repeat_region 16415..16679
/note="L2 repeat: matches 2286..2552 of consensus"
repeat_region 16741..16845
/note="L2 repeat: matches 1038..1144 of consensus"
misc_feature 16851..17272
/note="match: GSS: Em:A0057385"
repeat_region 16937..17018
/note="L2 repeat: matches 1158..1237 of consensus"
repeat_region 17197..17666
/note="L2 repeat: matches 2169..2696 of consensus"
repeat_region 18929..18970
/note="21 copies 2 mer ac 100% conserved"
misc_feature 19952..20346
/note="match: GSS: Em:A0384380"
repeat_region 20206..20305
/note="MIR repeat: matches 29..135 of consensus"
repeat_region 20669..20770
/note="MIR repeat: matches 47..152 of consensus"
repeat_region 20958..22168
/note="LIP repeat: matches 3882..5092 of consensus"
repeat_region 22177..23069
/note="LIPB3 repeat: matches 5254..6146 of consensus"
repeat_region 23070..23109
/note="20 copies 2 mer aa 80% conserved"
repeat_region 23111..23358
/note="LIM4 repeat: matches 2275..2185 of consensus"
repeat_region 23364..23429
/note="AluJ/FRAM repeat: matches 216..281 of consensus"
repeat_region 23431..25705
/note="LIM4 repeat: matches 2215..4526 of consensus"
repeat_region 25714..25910
/note="MER1B repeat: matches 141..337 of consensus"
repeat_region 25899..25948
/note="MTR1D repeat: matches 55..99 of consensus"
repeat_region 26064..26105
/note="21 copies 2 mer aa 76% conserved"
repeat_region 26107..26180
/note="MTR1D repeat: matches 353..429 of consensus"
repeat_region 26179..26319
/note="MER1B repeat: matches 2..139 of consensus"
repeat_region 26323..26824
/note="LIM4 repeat: matches 4529..5009 of consensus"
repeat_region 26825..27073
/note="MTR1D repeat: matches 240..505 of consensus"
repeat_region 27126..27177
/note="MTR1D repeat: matches 48..99 of consensus"
repeat_region 27447..27895
/note="LIMD3 repeat: matches 5913..6401 of consensus"
repeat_region 27910..28297
/note="LIPB10 repeat: matches 5767..6159 of consensus"
repeat_region 28313..34723
/note="LIPB7 repeat: matches 2..6143 of consensus"
repeat_region 34727..35016
/note="LIPB4 repeat: matches 5492..5781 of consensus"
repeat_region 35017..35281
/note="AluY repeat: matches 35..294 of consensus"
repeat_region 35282..36935
/note="LIP4 repeat: matches 3825..5493 of consensus"
repeat_region 36938..38320
/note="LIMC3 repeat: matches 6389..7735 of consensus"
repeat_region 39478..39533
/note="MER5B repeat: matches 96..148 of consensus"
repeat_region 39756..39872
/note="MIR repeat: matches 18..140 of consensus"
repeat_region 40558..40687
/note="MIR repeat: matches 35..152 of consensus"
repeat_region 41054..41249
/note="MIR repeat: matches 24..233 of consensus"
repeat_region 41306..41596

```

```

/note="AluX repeat: matches 1..298 of consensus"
repeat_region 41620..41731
/note="L2 repeat: matches 2571..2691 of consensus"
repeat_region 42492..42997
/note="MTR1C repeat: matches 1..464 of consensus"
repeat_region 42998..43171
/note="L2 repeat: matches 2075..2236 of consensus"
repeat_region 44118..44845
/note="LIMB3 repeat: matches 5447..6182 of consensus"
repeat_region 45984..46490
/note="LIR40a repeat: matches 2..519 of consensus"
repeat_region 47015..47093
/note="MER91A repeat: matches 105..186 of consensus"
repeat_region 47500..47917
/note="L2 repeat: matches 2309..2746 of consensus"
repeat_region 49126..49349
/note="MER20 repeat: matches 1..216 of consensus"
repeat_region 49611..49655
/note="MIR repeat: matches 219..262 of consensus"
repeat_region 49662..50194
/note="L2 repeat: matches 2096..2661 of consensus"
repeat_region 51919..52224
/note="AluSg repeat: matches 1..306 of consensus"
repeat_region 53562..53818
/note="LIMB8 repeat: matches 5819..6167 of consensus"
repeat_region 53817..54341
/note="LIMB6 repeat: matches 5614..6140 of consensus"
repeat_region 54343..54633
/note="AluSg repeat: matches 1..291 of consensus"
repeat_region 56076..56128
/note="L2 repeat: matches 2662..2709 of consensus"
repeat_region 56797..59200
/note="match: GSS: Em:A0146681"
repeat_region 58311..58743
/note="match: GSS: Em:A0890399"
repeat_region 58927..59200
/note="L2 repeat: matches 1543..1831 of consensus"
repeat_region 59422..59806
/note="LIM4 repeat: matches 472..862 of consensus"
repeat_region 59871..60217
/note="LIM4 repeat: matches 867..1199 of consensus"
repeat_region 60292..60367
/note="LIM4 repeat: matches 1204..1281 of consensus"

Query Match 70.7% Score 21.2 DB 9 Length 198917
Best Local Similarity 88.5% Pred. No. 30
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 ACGAGCGCTTGCATCTACCTTC 29
Db 187890 ACTGAGCTTTGAGATCTACCTTC 187915

RESULT 4
AF408636
LOCUS
DEFINITION Bacteriophage phi-12 P14 (14), P15 (15), core protein P7 (7),
RNA-dependent RNA polymerase P2 (2), NTPase P4 (4), and major core
structural protein P1 (1) genes, complete cds.
ACCESSION AF408636
VERSION AF408636
KEYWORDS AF408636.1 GI:15488102
SOURCE bacteriophage phi-12.
ORGANISM bacteriophage phi-12
REFERENCE 1 (bases 1 to 6751)

```

AUTHORS	Gottlieb, P.J., Potgieter, C., Wei, H. and Toporovsky, I.
TITLE	Characterization of Bacteriophage phi12
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 6751)
AUTHORS	Gottlieb, P.J., Potgieter, C., Wei, H. and Toporovsky, I.
TITLE	Direct Submission
JOURNAL	Submitted (10-ATG-2001)
FEATURES	Microbiology and Immunology, Sophie Davis School of Biomedical Education, City College of New York, 138th Street and Convent Avenue, New York, NY 10031, USA Location/Qualifiers

```

/organism="bacteriophage phi-12"
/db_xref="taxon:161736"
410..709
/gene="14"
410..709
CDS

```

gene
714..10163
/translation="NRKGRKTHNOKPDIIYTSDVVNRRTNRYRNVMSKOYWEDECK
PALDNEFKCYFVIGGASAILRLVWGJDSKTEDEVHLAEEMGNESSEFLVIKE"
/protein_id="AL01104.1"
/db_xref="GI:15488103"

```

/protein_id="AA101105.1"
/db_xref="GI:15488104"
/translation="MEGSDENGAVHTQINPSYKSGSEVRSKTYLFTPRDGA
IKRRIPIGTAGNAAILLCMDGSEVFRMDIDHAPRGKIKIINPDNTSHYELG-
1019..1528
/gene="7"
CDS
1019..1528

```

gene
CDS
1512..3491
/gene="2"
1512..3491
7Ctranslathion=MDFTDMSKNRLQRLKQVETSILMWSHNGVPIPTGFNNM
RTYTMIDLKVELPAVILGDPAODIALVPCARPLAVGNVITIRVDDPGQPSALVDA
EIGALHEVTSVNVLNDIADDEQLVKRNMISALELSETRKQEARQARDGIVRPLAVED
HGAIIDADE"

```

gene      3488..4483
          /gene="4"
CDS       3488..4483
          /gene="4"

```

```

/codon_start=1
/transl_table=11
/product="nrpase.p4"
/protein_id="AA01108.1"
/db_xref="GI:15488107"
/translation="MHIHYDKASFAKLRAAQAFAFHTDAPGSMFDTSGLESEVDGCT
PLAIVGESGDAIVDFDKNAORIVAKREKSVRAEDSVVQVENGFMKQGRGLYDI
TELVEGSEPVVAEFGHRAQSMVIVTKEKNGSKGPAIVHAIHGALGAGDKRAIYFEE
PLSGYNTDFENFVDDIARAMLQHRVITVDSLKNTYGAGAGTGGGSGISRGAFDILSDI
GMAASRGCVVYIASINPTNSNDKRIEVLKESRSNSTSLVISTDYGWCVLTITGSGG
LQRLHTLQTSVGEHSVLIHTHSQSGKQASGALQTVLNNDESLVLRILTSN"
/gene="1"
/497..668
/gene="1"
/497..668
/gene="1"
/codon_start=1
/transl_table=11
/product="major core structural protein p1"
/protein_id="AA01109.1"
/db_xref="GI:15488108"
/translation="MARKKIDDKRQMANRRPREFPNTSNNSSLATPAMANSVIENTI
VDRASGVVVEGCEGTEYNLSVVAHTIKVEGVENAKWFLERIVVKVLEANTIASNV
SIGEAGTADSLDLAAQNAVRAVEIVANVRNHEGKMSKLAIARAEGEALVAALGIEHRSN
TVSTADTSLDLAAQNAVRAVEIVANVRNHEGKMSKLAIARAEGEALVAALGIEHRSN
VOHIFDIYVAKIHVKLTLAASADMGVDSRMTLHPIDVEMSNRTYFVSAALDMPTGG
DGLKANDIHLEKAPLAIAISLKGSRNRYNVGDETELRYGKKTIVNVEGPEVFTGY
RMSLSVACQSVFENDAIIMPQKQANVPDQIDTEFVASMPKGNPSCGVPHNELI
NATHTLAETDIPRSFEYNGEGSAYAKYIGAPGFIISDVGDNFSEKELVLAAMDACYLMT
GPEDEEMVISTHTEFVNDILEDKRTVEKREKSPNRYVCAETISGAVYITTEKDIAPAEPR
AOLISDALETRVLDEEDKRTVEKREKSPNRYVCAETISGAVYITTEKDIAPAEPR
FVYPVHKRLVMOTIIAIIHSEFNELIAAKKVASSEPPGEGEDTDLDSNIAVTPOLVSF
MNSRSRAILITFAQSVPOYROIIVASVIMMRAYANRKSADIDILKGTMOOOSFANAYD
VLGLVLTINGLSGSEFIDILEDASVMAIMLIGSDRVKN"

```

Query Match	70.0%	Score 21	DB 7	Length 6751
Best Local Similarity	82.8%	Pred. No. 37		
Matches 24	Conservative 0	Mismatches	5	Indels 0
				Gaps 0
QY	1	TGACTCAGTCCCTTGATCATCTACCTTC 29		
Db	4449	TGATATCGTCTCTTCGTCTCATCTCC 4477		
RESULT 5				
AC013516/c				
LOCUS	AC013516	30265:bp	DNA	linear
DEFINITION	Homo sapiens clone RP11-115K5, LOW-PRESS SEQUENCE SAMPLING.			HTG 13-JUL-2000
AC013516				
AC013516.2	GI:9112504			
HTG, HTGS, PHASED.				
SOURCE	Homo sapiens.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
TITLE	1 (bases 1 to 30265)			
JOURNAL	Bitren,B., Linton,L., Nusbaum,C. and Lander,E.			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 30265)			
	Bitren,B., Linton,L., Nusbaum,C.			

Baldwin, J., Barua, N., Beckerly, R., Boguslavsky, L., Boukhalter, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collimore, A., Cooke, P., Deastellano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, J., Haggos, B., Hearford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lebockzy, J., Lieu, C., Locke, K., MacDonald, P., Marguis, N., McEwan, P., McGuirk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Sudriaman, A., Talamas, J.,

TITLE JOURNAL COMMENT

Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6403765.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

Project information

Center project name: 115_K_5

NOTE: This record contains 34 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

1 753: contig of 753 bp in length
1 754 853: gap of 100 bp
1 854 1682: contig of 829 bp in length
1 1683 1782: gap of 100 bp
1 1783 2620: contig of 838 bp in length
1 2621 2720: gap of 100 bp
1 2721 3546: contig of 826 bp in length
1 3547 3646: gap of 100 bp
1 3647 4551: contig of 905 bp in length
1 4552 4651: gap of 100 bp
1 4652 5461: contig of 810 bp in length
1 5462 5561: gap of 100 bp
1 5562 6319: contig of 758 bp in length
1 6320 6419: gap of 100 bp
1 6420 7229: contig of 810 bp in length
1 7230 7329: gap of 100 bp
1 7330 8105: contig of 776 bp in length
1 8106 8205: gap of 100 bp
1 8206 9031: contig of 826 bp in length
1 9032 9131: gap of 100 bp
1 9132 9895: contig of 764 bp in length
1 9896 9995: gap of 100 bp
1 10791 10890: contig of 795 bp in length
1 10891 11726: contig of 836 bp in length
1 11727 11826: gap of 100 bp
1 11827 12605: contig of 779 bp in length
1 12606 12705: gap of 100 bp
1 12706 13525: contig of 820 bp in length
1 13526 13625: gap of 100 bp
1 13626 14374: contig of 749 bp in length
1 14375 14474: gap of 100 bp
1 14475 15188: contig of 714 bp in length
1 15189 15288: gap of 100 bp
1 15289 16131: contig of 843 bp in length
1 16132 16231: gap of 100 bp
1 16232 16933: contig of 702 bp in length
1 16934 17033: gap of 100 bp
1 17034 17842: contig of 809 bp in length
1 17843 17942: gap of 100 bp
1 17943 18731: contig of 789 bp in length
1 18732 18831: gap of 100 bp
1 18832 19608: contig of 777 bp in length
1 19609 19708: gap of 100 bp

```

```

19709 20487: contig of 779 bp in length
* 20488 20587: gap of 100 bp
* 20588 21395: contig of 808 bp in length
* 21396 21495: gap of 100 bp
* 21496 22261: contig of 766 bp in length
* 22262 22361: gap of 100 bp
* 22362 23158: contig of 793 bp in length
* 23159 23258: gap of 100 bp
* 23259 24052: contig of 794 bp in length
* 24053 24152: gap of 100 bp
* 24153 24927: contig of 775 bp in length
* 24928 25027: gap of 100 bp
* 25028 25812: contig of 783 bp in length
* 25813 25912: gap of 100 bp
* 25913 26717: contig of 805 bp in length
* 26718 26817: gap of 100 bp
* 26818 27628: contig of 811 bp in length
* 27629 27728: gap of 100 bp
* 27729 28468: contig of 740 bp in length
* 28469 28568: gap of 100 bp
* 28569 29356: contig of 788 bp in length
* 29357 29456: gap of 100 bp
* 29457 30265: contig of 809 bp in length.
Location/Qualifiers
1..30265
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-115K5"
/clone_11b="RP11-11 Human Male BAC"

```

BASE COUNT 7067 a 6735 c 6224 g 6406 t 3833 others

Query Match 70.0% Score 21; DB 2; Length 30265;
Best Local Similarity 82.8% Pred No. 37;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

QY 2 GGACTCAGTCTGCTGTCATCTCACCCTTC 30
DB 327 GGACTCAGTCTGCTGTCATCTCACCCTTC 259

```

RESULT 6
HS32110 93312 bp DNA linear PRI 12-DEC-1999
LOCUS Human DNA sequence from clone RPI-32110 on chromosome 22 Contains a
DEFINITION pseudogene similar to TCEB1L (transcription elongation factor B
(SIL1), polypeptide 1-like), a CA repeat (D22S1159), STRs and
GSSs, n, complete sequence.

ACCESSION

285994.1 GI:2342582

VERSION HTG: D22S1159; repeat polymorphism: TCEB1L.

SOURCE

Homo sapiens.

Organism Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 93312)

REFERENCE

1 (bases 1 to 93312)

AUTHORS

Ho, S.

JOURNAL

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

COMMENT

On Aug 24, 1997 this sequence version replaced gi:1841910.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows: An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'

Feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at: http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/Map/Chr22> RPI-32110 is from the library RPI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://daccpac.med.buffalo.edu/VECTOR> PCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RPI-32110. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The true left end of clone RPI-32110 is at 1 in this sequence.

FEATURES

SOURCE

1. 93312
location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="22"

/clone_id="RPI-32110"

/clone_id="RPI-1"

25. 196
/note="match: STS: Em:HS32110T"

671. 810
/note="35 copies 4 mer ggtg 70 conserved"

1041. 1091
/note="L2 repeat: matches 2688. 2746 of consensus"

1051. 1126
/note="MIR repeat: matches 174. 256 of consensus"

1454. 1608
/note="L2 repeat: matches 1986. 2144 of consensus"

1609. 1904
/note="AluSg repeat: matches 1. 308 of consensus"

1905. 1959
/note="L2 repeat: matches 2144. 2197 of consensus"

2224. 2389
/note="MIR repeat: matches 89. 262 of consensus"

2471. 2550
/note="L2 repeat: matches 2667. 2745 of consensus"

3038. 3106
/note="MIR repeat: matches 56. 128 of consensus"

3294. 3530
/note="MIR repeat: matches 13. 258 of consensus"

4209. 4325
/note="MIR repeat: matches 48. 175 of consensus"

6405. 6487
/note="MIR repeat: matches 46. 131 of consensus"

6608. 6910
/note="AluSg repeat: matches 1. 303 of consensus"

8780. 8899
/note="L2 repeat: matches 2091. 2218 of consensus"

8908. 8971
/note="L2 repeat: matches 2388. 2453 of consensus"

8972. 9266
/note="AluX repeat: matches 1. 296 of consensus"

9267. 9581
/note="L2 repeat: matches 2453. 2750 of consensus"

complement(9882..10117)
/note="match: STS: Em:Z53151; match: STS: Em:HSB0292H9"

9936. 9977
/note="21 copies 2 mer tg 95 conserved"

9939. 9978
/note="10 copies 4 mer ggtg 95 conserved"

10003. 10283
/note="AluSg repeat: matches 1. 280 of consensus"

10524. 10766
/note="L2 repeat: matches 2260. 2498 of consensus"

11167. 11346

repeat_region /note="MIR repeat: matches 1. 184 of consensus"

11347. 11661
/note="AluX repeat: matches 3. 306 of consensus"

11662. 11892
/note="MIR repeat: matches 184. 400 of consensus"

11893. 12203
/note="AluX repeat: matches 1. 310 of consensus"

12204. 12256
/note="MIR repeat: matches 400. 453 of consensus"

12368. 12523
/note="MIR repeat: matches 108. 262 of consensus"

12810. 12915
/note="L2 repeat: matches 2438. 2536 of consensus"

12915. 12987
/note="MIR repeat: matches 56. 121 of consensus"

13585. 13632
/note="L2 repeat: matches 2649. 2698 of consensus"

13826. 14242
/note="MIR repeat: matches 1. 466 of consensus"

14858. 14980
/note="MIR repeat: matches 56. 192 of consensus"

15124. 15405
/note="AluSg repeat: matches 1. 288 of consensus"

15489. 15585
/note="MIR repeat: matches 59. 148 of consensus"

15637. 15699
/note="L2 repeat: matches 2348. 2411 of consensus"

16503. 16575
/note="MIR repeat: matches 27. 104 of consensus"

16708. 17082
/note="MIR repeat: matches 1. 382 of consensus"

18010. 18204
/note="MIR repeat: matches 35. 239 of consensus"

18260. 18357
/note="MIR repeat: matches 42. 144 of consensus"

19043. 19101
/note="MIR repeat: matches 92. 152 of consensus"

19587. 19718
/note="MIR repeat: matches 25. 185 of consensus"

20352. 20528
/note="MIR repeat: matches 33. 212 of consensus"

20896. 21504
/note="MIR repeat: matches 3. 610 of consensus"

21540. 21702
/note="MIR repeat: matches 61. 233 of consensus"

22247. 22605
/note="MIR repeat: matches 193. 547 of consensus"

23244. 23555
/note="AluSg repeat: matches 1. 310 of consensus"

23729. 23843
/note="MIR repeat: matches 34. 144 of consensus"

24584. 24940
/note="L2 repeat: matches 2315. 2671 of consensus"

25020. 25155
/note="MIR repeat: matches 60. 188 of consensus"

25342. 25692
/note="L2 repeat: matches 2124. 2478 of consensus"

26105. 26211
/note="MIR repeat: matches 53. 160 of consensus"

26212. 26526
/note="AluX repeat: matches 1. 308 of consensus"

26527. 26832
/note="MIR repeat: matches 160. 459 of consensus"

27434. 27544
/note="MIR repeat: matches 80. 195 of consensus"

28231. 28542
/note="AluSg repeat: matches 1. 312 of consensus"

30366. 31108
/note="MIR repeat: matches 1027. 1841 of consensus"

31858. 33268
/note="LIM4 repeat: matches 2210. 3688 of consensus"

33136. 33471
/note="match: GSS: Em:AD617050"

misc_feature


```

repeat_region      33470. .33545  

                    /note="19 copies 4 mer cccct 80 conserved"  

repeat_region      33547. .33682  

                    /note="68 copies 2 mer ct 73 conserved"  

repeat_region      33686. .33750  

                    /note="MTR1D repeat: matches 182. .249 of consensus"  

repeat_region      34043. .35036  

                    /note="L1M4 repeat: matches 3723. .4744 of consensus"  

repeat_region      35209. .35381  

                    /note="MIR repeat: matches 23. .219 of consensus"  

repeat_region      36094. .36230  

                    /note="MIR repeat: matches 117. .260 of consensus"  

repeat_region      36714. .35893  

                    /note="MIR repeat: matches 79. .262 of consensus"  

                    37498. .37754  

                    /note="match: STR: Em:G49322"  

                    37579. .37934  

                    /note="match: STR: Em:G49323"  

repeat_region      37971. .38036  

                    /note="L2 repeat: matches 2127. .2194 of consensus"  

repeat_region      38072. .38180  

                    /note="MIR repeat: matches 16. .143 of consensus"  

repeat_region      38333. .38579  

                    /note="L2 repeat: matches 2358. .2671 of consensus"  

Query Match      70.0%; Score 21; DB 9; Length 93312;  

Best Local Similarity 82.8%; Pred. No. 37;  

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

RESULT 7	AC0004943/c	141313 bp	DNA	linear	PRI 22-MAY-2002
LOCUS	AC0004943				
DEFINITION	Homo sapiens PAC clone RP5-991G20 from 16, complete sequence.				
ACCESSION	AC0004943				
VERSION	AC0004943.2	GI:21070767			
KEYWORDS	HTG.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE	1 (bases 1 to 141313)				
AUTHORS	Suiston, J.E. and Waterston, R.				
TITLE	Toward a complete human genome sequence				
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)				
MEDLINE	99063792				
PUBMED	9847074				
REFERENCE	2 (bases 1 to 141313)				
AUTHORS	Kozlowicz, A., Laplant, Y. and Hottic, M.				
TITLE	The sequence of Homo sapiens PAC clone RP5-991G20				
JOURNAL	Unpublished (2001)				
REFERENCE	3 (bases 1 to 141313)				
AUTHORS	Waterston, R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA				
REFERENCE	4 (bases 1 to 141313)				
AUTHORS	Waterston, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-NOV-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA				
REFERENCE	5 (bases 1 to 141313)				
AUTHORS	Waterston, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA				
REFERENCE	6 (bases 1 to 141313)				
AUTHORS	Waterston, R.				

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (22-MAY-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 22, 2002 this sequence version replaced gi:3924671.

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: saplens@wustl.wustl.edu
----- Summary Statistics -----
Center project name: H_DJ0991620

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION: This clone was derived from human PAC library RPECI-5, prepared by Pieter de Jong and coworkers at <http://www.chori.org> using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor. The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong. VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is CTD-2532L16, 200 bp overlap
Actual start of this clone is at base position 1 of R55-991G20;
actual end is at base position 141313 of R55-991G20.

```
FEATURES
source
location/Qualifiers
1..141313
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16"
/clone="RP5-991G20"
/clone_1lb="RPCI-5"
85..235
/rpt_family="MIR"
1219..1292
/rpt_family="L2"
1386..1786
/note="similar to EST W87623 (NM:91401687) zh6sh07.r1"
repeat_region
2123..2228
/rpl_family="L2"
2352..2613
/rpt_family="Alu"
2743..2916
/rpt_family="L1"
2978..3329
/rpt_family="Alu"
3396..3517
/rpt_family="Alu"
3521..4092
/rpt_family="L1"
4285..4342
repeat_region
```

REHDSAEGEENGTPRELLPDILKAKKEKLAPGGSEPSMLPRILASDARGNATK
LENGFEIYOYNENKOKYORNGKTGDEENILEDCSCCKELSNLILIKSHHEH
ONAFPTPEKOLERRAKORYRHDIDLYPLRQTEPEPPPPPPEPPLPAAPPPOSTEP
PASAPFTSTITAPAPOSPVLTLQOONKRPRIEPLFSPILAMOTMQLTLPALPOLGPV
PIPADIAOLYQHOLNPRLLOONKRPRIEDDOLRYLOYFOINNSPSEJOIKEMAN
KSGLPOYTKHWFRTTLKERORNRKSDSYNPNPISLETIKIDSPPSPPEPKOE
WSKSRSTRIFNDYOLRYLODFPFANMYNDEDEREISNLINLTRIVVFONARQ
ARKYENOGEGEDGERRELTDYRTIRTSNYOCKRSLVFORFDLIKOKLYCRK
DEEEODOSONDMSDAEILLTPRSSCSIPMSOAYSAIPAPANNNTASAFLJOLA
AEBLATFNSTKGADGEEKLAENASBPNGIOEQKPELQOOOPDEQKWTMPQOL
LPOULSLPLPOPPOAPPPOCPLPOSSPSPSLSHLPLKPLHTSTPOOLANPLPPL
PYCCOCHKLARPSTEHWEOHOHLFISAOMFTHPOLDRLSDMLPMULDPSNPLLA
OLLSALTQIPIASSATSPTSPTSTMTNLKLEKASASGENDESGTGEEPORRK
KITTEPOLLETLKYKYLDSNPTKMKMDLHAHVGLKARYVOGWONTARAERKQCF
AVGPAGARRCPFCRALFKAKTALEAHIRRHMHKRAKNYTATSMILDCGGLQ
RGDLFDGTSEFHLPSSSDCGVPLVSYSKTMELSPRTLISBSTIVEIEEFESPSL
SVVNINFQRTKLDNDCCSVNVAITDTTTDBCNANDSATGAIEETKSSAPNBGL

70.0%; Score 21; DB 9; Length 141313;
Best Local Similarity 82.8%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0.

Query Match

Db 77445 GGACTGACCTCTTAGTCAGCTCCCACT 77417

OY 2 GGACTGACCTCTTAGTCAGCTCCCTCT 30
||||| | | | | | | | | | | | | | |
AC108030 Homo sapiens BAC clone Rp11-100C9 from 4, complete sequence.
AC108030 AC009865
AC108030.4 GI:19551171
HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 147091)
Sultson,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
JOURNAL MEDLINE
PUBMED 99063792
9847074
2 (bases 1 to 147091)
Isak,A., Haekenson,W. and Creason,K.
The sequence of Homo sapiens BAC clone Rp11-100C9
Unpublished (2001)
3 (bases 1 to 147091)
Waterston,R.H.
Direct Submission
Submitted (24-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108 USA
4 (bases 1 to 147091)
Waterston,R.H.
Direct Submission
Submitted (13-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 147091)
Waterston,R.
Direct Submission
Submitted (20-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 20, 2002 this sequence version replaced gi:19587786.
-- Genome Center
Center: Washington University Genome Sequencing Center
Accession code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplenetw@wustl.edu
Summary Statistics
Center project name: H.NH0100C09

REFERENCE AUTHORS TITLE JOURNAL COMMENT

Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-384K6; the clone sequenced to the right is RP11-358D17, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-100C9; actual end is at base position 34330 of RP11-358D17.

The sequence of AC009865 has been incorporated into AC108030.

FEATURES

source

```
1. 147091
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-100C9"
/clone_idb="RPCI-11"
76. 372
/rpl_family="Alu"
misc_feature
2787. 2814
/note="match to EST B1048881 (NID:g14455503)"
misc_feature
3137. 3197
/note="match to EST AM805082 (NID:g7856965)"
repeat_region
3269. 3554
/rpl_family="AT-rich"
repeat_region
4079. 4181
/rpl_family="Alu"
misc_feature
4082. 4181
/note="similar to EST BM461794 (NID:g18510834)"
misc_feature
4082. 4181
/note="match to EST AU130547 (NID:g10990901)"
misc_feature
4788. 4980
/note="match to EST BG687120 (NID:g13918517)"
misc_feature
4788. 4980
/note="similar to Mus musculus EST BF302831 (NID:g11249388)"
misc_feature
4789. 4980
/note="match to EST AU130547 (NID:g10990901)"
misc_feature
4789. 4980
/note="match to EST BG687120 (NID:g13918517)"
misc_feature
4789. 4980
/note="similar to EST BM461794 (NID:g18510834)"
repeat_region
6760. 6907
```

```
misc_feature
9813. 9937
/rpl_family="MERL_type"
misc_feature
9813. 9937
/note="match to EST AM805082 (NID:g7856965)"
misc_feature
9813. 9937
/note="similar to EST BM461794 (NID:g18510834)"
misc_feature
9813. 9937
/note="similar to Mus musculus EST BF302831 (NID:g11249388)"
misc_feature
9813. 9886
/note="match to EST BG687120 (NID:g13918517)"
misc_feature
9813. 9834
/note="match to EST AU130547 (NID:g10990901)"
repeat_region
10777. 10816
/rpl_family="AT-rich"
repeat_region
11253. 11548
/rpl_family="L2"
repeat_region
12246. 12801
/rpl_family="L1"
misc_feature
12483. 12711
/note="match to EST AM805082 (NID:g7856965)"
repeat_region
12802. 13108
/rpl_family="Alu"
repeat_region
13109. 13147
/rpl_family="L1"
repeat_region
13299. 13356
/rpl_family="L2"
repeat_region
13357. 13390
/rpl_family="(TGA)n"
misc_feature
13489. 13687
/note="similar to Homo sapiens EST B1048935 (NID:g14455557)"
misc_feature
13497. 13826
/note="match to EST BF910439 (NID:g12301897)"
repeat_region
13639. 13745
/rpl_family="MERL_type"
repeat_region
13915. 14163
/rpl_family="L2"
repeat_region
14366. 14604
/rpl_family="MIR"
repeat_region
16289. 16314
/rpl_family="AT-rich"
misc_feature
16878. 16879
/note="match to EST B1712490 (NID:g15688185) le08e05.x1"
misc_feature
16889. 17009
/note="match to EST BF031310 (NID:g10739022)"
misc_feature
16894. 17009
/note="similar to EST BM461794 (NID:g18510834)"
misc_feature
16894. 17009
/note="similar to Mus musculus EST BF302831 (NID:g11249388)"
misc_feature
17068. 17577
/note="match to EST B1712490 (NID:g15688185) le08e05.x1"
misc_feature
17411. 17934
/note="match to EST B1712728 (NID:g15688423) le08e05.y1"
misc_feature
18993. 19340
/note="match to EST BF841853 (NID:g12195175)"
misc_feature
20265. 20409
/note="similar to Mus musculus EST BF302831 (NID:g11249388)"
misc_feature
20265. 20403
/note="match to EST BF031310 (NID:g10739022)"
misc_feature
20274. 20403
/note="similar to EST BM461794 (NID:g18510834)"
misc_feature
20311. 20403
/note="match to EST BE247482 (NID:g9099239)"
misc_feature
20316. 20403
/note="match to EST AM96773 (NID:g8257007)"
misc_feature
20350. 20403
/note="match to EST AW369711 (NID:g6874365)"
misc_feature
20487. 20786
/note="match to EST BE831699 (NID:g10264077)"
repeat_region
/rpl_family="Alu"
```

```

misc_feature      21804..22199
                  /note="match to EST AV707981 (NID:g10725246)"
repeat_region     21946..22011
                  /rpt_family="MIR"
repeat_region     26140..26182
                  /rpt_family="AT-rich"
repeat_region     26650..27046
                  /rpt_family="L2"
misc_feature      26984..27438
                  /note="match to EST BF059079 (NID:g10812975)"

Query Match      70.0% Score 21; DB 9; Length 147091;
Best Local Similarity 82.8% Pred. No. 37;
Matches 124; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db      51688 GGGCTCAGTACTGTGTCATCTCTCTCT 51660
        ||||||| ||||||| |||||||
OY      2 GGACTCAGTCTTGTCATCTCACTCTCT 30
        ||||||| ||||||| |||||||

RESULT 9
AC024132      151552 bp DNA linear PRI 07-NOV-2001
LOCUS      Homo sapiens BAC clone RP11-415C15 from 4, complete sequence.
ACCESSION      AC024132
VERSION      AC024132.7 GI:14589728
KEYWORDS      HTG.
SOURCE      Homo sapiens.
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 151552)
AUTHORS      Sulston,J.E. and Waterston,R.
JOURNAL      Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE      98063792
PUBMED      9847074
REFERENCE      2 (bases 1 to 151552)
AUTHORS      Goya,E. and Abbott,A.
JOURNAL      The sequence of Homo sapiens BAC clone RP11-415C15
              Unpublished
REFERENCE      3 (bases 1 to 151552)
AUTHORS      Waterston,R.H.
JOURNAL      Direct Submission
              Submitted (24-FEB-2000) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
              4 (bases 1 to 151552)
              Waterston,R.H.
              Direct Submission
              Submitted (03-JUL-2001) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
              5 (bases 1 to 151552)
              Waterston,R.
              Direct Submission
              Submitted (07-NOV-2001) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              On Jul 3, 2001 this sequence version replaced g1:14151042.
              Genome Center
              Center: Washington University Genome Sequencing Center
              Web site: http://genome.wustl.edu/gsc
              Contact: saplens@wustl.wustl.edu
              Summary Statistics
              Center project name: H_NH0415C15

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPI1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateo,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is AC006568. Actual start of this clone is at base position 1 of RP11-415C15.

The region between 36765 to 36883 is covered only by a single

FEATURES

source

```

repeat_region     1..151552
                  /location/Qualifiers
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="4"
                  /map="4"
                  /clone="RP11-415C15"
                  /clone_id="RPI1-11"
                  2..947
                  /rpt_family="L1"
repeat_region     988..1084
                  /rpt_family="L1"
repeat_region     2351..2559
                  /rpt_family="MaLR"
repeat_region     2745..3273
                  /rpt_family="MERL-type"
repeat_region     3625..3931
                  /rpt_family="Alu"
repeat_region     4992..5248
                  /rpt_family="Alu"
repeat_region     5470..5776
                  /rpt_family="Alu"
repeat_region     5629..5658
                  /rpt_family="Alu"
repeat_region     6330..6682
                  /rpt_family="(T)n"
repeat_region     6999..7352
                  /rpt_family="ERV1"
repeat_region     7350..7379
                  /rpt_family="MaLR"
repeat_region     7353..7660
                  /rpt_family="AT-rich"
repeat_region     7661..7769
                  /rpt_family="Alu"
repeat_region     7853..7878
                  /rpt_family="MaLR"
repeat_region     8615..8645
                  /rpt_family="(T)n"
repeat_region     8619..8929
                  /rpt_family="(TTTC)n"
repeat_region     9339..9651
                  /rpt_family="Alu"
repeat_region     9627..9651
                  /rpt_family="Alu"

```

```

repeat_region      /rpl_family="(A)n"
10012. 10093
repeat_region      /rpl_family="MERL_type"
11138. 11326
repeat_region      /rpl_family="MERL_type"
11397. 11758
repeat_region      /rpl_family="MALR"
12010. 12337
repeat_region      /rpl_family="Alu"
12315. 12337
repeat_region      /rpl_family="(GAAA)n"
12522. 12700
repeat_region      /rpl_family="CR1"
13352. 13648
repeat_region      /rpl_family="Alu"
13661. 13790
repeat_region      /rpl_family="MALR"
13791. 13972
repeat_region      /rpl_family="MERL_type"
13973. 14064
repeat_region      /rpl_family="MALR"
14300. 14613
repeat_region      /rpl_family="MIR"
15241. 16178
repeat_region      /rpl_family="L1"
16173. 16901
repeat_region      /rpl_family="L1"
17238. 17666
repeat_region      /rpl_family="L1"
18792. 18885
repeat_region      /rpl_family="L2"
19121. 19301
repeat_region      /rpl_family="L2"
19294. 19319
repeat_region      /rpl_family="(TGA)n"
19954. 20186
repeat_region      /rpl_family="ERVU"
21116. 21231
repeat_region      /rpl_family="MALR"
21289. 21485
repeat_region      /rpl_family="CR1"
22221. 22536
repeat_region      /rpl_family="Alu"
22509. 22536
repeat_region      /rpl_family="(A)n"
22595. 22802
repeat_region      /rpl_family="L1"
22830. 22940
repeat_region      /rpl_family="L1"
22941. 23305
repeat_region      /rpl_family="MALR"
23306. 23393
repeat_region      /rpl_family="L1"
23407. 23698
repeat_region      /rpl_family="Alu"
23700. 23809
repeat_region      /rpl_family="(TA)n"
23813. 24012
repeat_region      /rpl_family="L1"
24045. 24386
repeat_region      /rpl_family="L2"
24412. 24522
repeat_region      /rpl_family="ERVU"
24602. 24728
repeat_region      /rpl_family="ERVU"
26104. 26128
repeat_region      /rpl_family="AT-rich"
26508. 26576
repeat_region      /rpl_family="MALR"
26712. 27121
repeat_region      /rpl_family="MALR"
27122. 27419
repeat_region      /rpl_family="Alu"

```

```

repeat_region      27420. 27680
repeat_region      /rpl_family="MALR"
27682. 27951

Query Match      70.0%; Score 21; DB 9; Length 151552;
Best Local Similarity 82.8%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Cy      2 GGACTGACCTCTGTGTCATCTCACCTTCT 30
Db      112885 GGCTGACCTCTGTGTCATCTCTCTCTCT 112913

RESULT 10
AC121186/c
LOCUS      156128 bp DNA linear HTG 23-JUL-2002
DEFINITION Rattus norvegicus clone CH230-331A2, *** SEQUENCING IN PROGRESS
ACCESSION AC121186
VERSION AC121186.2 GI:21902677
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 156128)
Muzny,D.M., Adams,C., Adlo-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaraltinge,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbaria,J., Benton,J., Blumage,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.I., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,U., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devalla,M.L., Davys,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,U., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,D.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolliver,S., Joudah,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mathoney,E., McLeod,M.P., Meador,M., Mel,G., Metker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,S.,
Nguyen,N., Nickerson,E., Nwokoko,S., Ogulu,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojudoxan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoochari,N., Sisson,I.,
Sodergren,E., Somaite,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Swalek,A., Tabor,P., Tameris,A., Tameris,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usman,I., Vasquez,L., Vera,V., Villalobos,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Woodden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE      Unpublished
JOURNAL    2 (bases 1 to 156128)
REFERENCE Worley,K.C.
AUTHORS    Submitted (16-MAY-2002) Human Genome Sequencing Center, Department
TITLE      JOURNAL
JOURNAL    Direct Submission
AUTHORS     Direct Submission
JOURNAL    Submitted (16-MAY-2002) Human Genome Sequencing Center, Department

```

REFERENCE
AUTHORS:
TITLE
JOURNAL

COMMENT

Of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
(bases 1 to 156128)
Mortley, K.C.
Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 18, 2002 this sequence version replaced g1:20806215.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GMP

Center clone name: CH230-331A2

Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap: version 0.990329

Consensus quality: 104035 bases at least Q40

Consensus quality: 110796 bases at least Q30

Consensus quality: 115903 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 57 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1029: contig of 1029 bp in length
1030 1129: gap of unknown length
1130 2199: contig of 1070 bp in length
2200 2299: gap of unknown length
2300 3670: contig of 1371 bp in length
3671 3770: gap of unknown length
3771 5275: contig of 1505 bp in length
5276 5375: gap of unknown length
5376 6409: contig of 1034 bp in length
6410 6509: gap of unknown length
6510 7565: contig of 1056 bp in length
7566 7665: gap of unknown length
7666 8816: contig of 1151 bp in length
8817 8916: gap of unknown length
8917 10099: contig of 1183 bp in length
10100 10199: gap of unknown length
10200 11446: contig of 1247 bp in length
11447 12748: gap of unknown length
12749 12848: contig of 1202 bp in length
12849 14095: gap of unknown length
14096 14194: contig of 1246 bp in length
14195 15241: gap of unknown length
15242 15341: contig of 1047 bp in length
15342 16738: gap of unknown length
16739 16838: contig of 1397 bp in length
16839 18428: gap of unknown length
18429 18528: contig of 1590 bp in length
18529 19576: gap of unknown length
19577 19676: contig of 1048 bp in length
19678 21464: gap of unknown length
21465 21564: contig of 1788 bp in length
21565 23696: gap of unknown length
23697 23796: contig of 2132 bp in length
23797 25356: gap of unknown length
25357 25656: contig of 1760 bp in length
25657 26780: gap of unknown length
26781 26880: contig of 1124 bp in length
26880: gap of unknown length

26881 28222: contig of 1342 bp in length
28223 28322: gap of unknown length
28323 29760: contig of 1438 bp in length
29761 29860: gap of unknown length
29861 31634: contig of 1774 bp in length
31635 31735: gap of unknown length
31736 33237: contig of 1502 bp in length
33237 33387: gap of unknown length
33387 35288: contig of 1951 bp in length
35288 35388: gap of unknown length
35388 38093: contig of 2706 bp in length
38093 38194: gap of unknown length
38194 40557: contig of 2364 bp in length
40558 40657: gap of unknown length
40658 44259: contig of 3602 bp in length
44260 44359: gap of unknown length
44360 46117: contig of 1757 bp in length
46117 46217: gap of unknown length
46217 48048: contig of 1831 bp in length
48048 48148: gap of unknown length
48148 50033: contig of 1885 bp in length
50033 50133: gap of unknown length
50133 52336: contig of 2203 bp in length
52336 52436: gap of unknown length
52436 54503: contig of 2067 bp in length
54503 54603: gap of unknown length
54603 57315: contig of 2713 bp in length
57315 57415: gap of unknown length
57416 59726: contig of 2311 bp in length
59727 59826: gap of unknown length
59827 63021: contig of 3195 bp in length
63021 63121: gap of unknown length
63121 65230: contig of 2109 bp in length
65230 65330: gap of unknown length
65330 67944: contig of 2614 bp in length
67944 68045: gap of unknown length
68045 70633: contig of 2589 bp in length
70634 70733: gap of unknown length
70734 73063: contig of 2330 bp in length
73064 73163: gap of unknown length
73164 76481: contig of 3318 bp in length
76482 76581: gap of unknown length
76582 78389: contig of 1808 bp in length
78390 78489: gap of unknown length
78490 82943: contig of 4454 bp in length
82944 83044: gap of unknown length
83044 86661: contig of 3618 bp in length
86662 86761: gap of unknown length
86762 89911: contig of 3150 bp in length
89912 90011: gap of unknown length
90012 94436: contig of 4425 bp in length
94437 94537: gap of unknown length
94537 98167: contig of 3631 bp in length
98168 98267: gap of unknown length
98268 102900: contig of 4633 bp in length
102901 103000: gap of unknown length
103001 107581: contig of 4581 bp in length
107582 107681: gap of unknown length
107682 110850: contig of 3169 bp in length
110851 110950: gap of unknown length
110951 113470: contig of 2520 bp in length
113471 113570: gap of unknown length
113571 117341: contig of 3771 bp in length
117342 117441: gap of unknown length
117442 121926: contig of 4485 bp in length
121927 122026: gap of unknown length
122027 124628: contig of 2602 bp in length
124629 124728: gap of unknown length
124729 132125: contig of 7397 bp in length

Query Match 70.0%; Score 21; DB 2; Length 156128;
Best Local Similarity 82.8%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 GGACGAGTCCTGGTCATCTCACCCTTCT 30
 DB 34085 GGATCAGGCGCTGGCTCTCTCACCCTACT 34057

RESULT 11
 AC090937 160696 bp DNA linear PRI 19-MAR-2001
 LOCUS Homo sapiens chromosome 3 clone RP11-316A10 map 3p, complete
 DEFINITION sequence.
 AC090937 AC012036
 VERSION AC090937.1 GI:13378165
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS Zhang, H., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C.,
 Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q.,
 Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N.,
 Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y.,
 Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q.,
 Qi, X., Song, L., Song, S., Sun, M., Sun, Y., Tan, X., Tao, R.,
 Wang, H., Wang, J., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R.,
 Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y.,
 Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, L.,
 Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y.,
 Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.
 Chromosome 3p genomic sequence
 Unpublished

TITLE
 JOURNAL 2 (bases 1 to 160696)
 REFERENCE
 AUTHORS Zhang, H., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C.,
 Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q.,
 Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N.,
 Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y.,
 Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q.,
 Qi, X., Song, L., Song, S., Sun, M., Sun, Y., Tan, X., Tao, R.,
 Wang, H., Wang, J., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R.,
 Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y.,
 Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, L.,
 Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y.,
 Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.
 Direct Submission
 Submitted (19-MAR-2001) Human Genomic Center, Institute of
 Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
 100101, P.R.China
 On Mar 19, 2001 this sequence version replaced gi:9719566.

COMMENT
 -----Genome Center
 Center: Beijing Center
 Website: http://hg.19t.ac.cn
 http://www.genomics.org.cn
 Contact: hgc@19t.ac.cn
 -----Project Information
 Center project name: RP11-316A10
 -----Summary Statistics
 Sequencing vector: pUC18; 100% of reads
 Chemistry: Dye-terminator; BT 5% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 744 bases at least Q40
 Consensus quality: 974 bases at least Q30
 Consensus quality: 1088 bases at least Q20
 Insert size: 1070; sum-of-contigs
 Quality coverage: 1.20x in Q20 bases; sum-of-contigs

FEATURES
 source location/Qualifiers
 1..160696
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 41995 a 35161 c 35897 g 47643 t
 ORIGIN
 Query Match 70.0%; Score 21; DB 9; Length 16066;
 Best Local Similarity 82.8%; Pred. No. 37;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 GGACGAGTCCTGGTCATCTCACCCTTCT 30
 DB 89672 GGAGTCAGTCCTGGCTCTCTCACCCTACT 89700

RESULT 13
 AC025163 177876 bp DNA linear HTG 26-JUN-2001
 LOCUS Homo sapiens chromosome 3 clone RP11-533F2, WORKING DRAFT SEQUENCE,
 DEFINITION 17 unordered pieces.
 AC025163
 VERSION AC025163.17 GI:14547360

REFERENCE
 AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 161604)
 Homo sapiens 161,604 genomic DNA of 11q
 Published Only in Database (2000)

TITLE
 JOURNAL 2 (bases 1 to 161604)
 REFERENCE
 AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission
 Submitted (02-MAY-2000) Masahiro Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou, Tsukuba, Ibaraki, Japan
 (E-mail: hattori@gsc.riken.go.jp, URL: http://dnp.gsc.riken.go.jp/,
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 On Jan 22, 2001 this sequence version replaced gi:9927284.

COMMENT
 STS Markers (ePCR):
 30042 30260 D1S3059 Homo sapiens
 32655 32849 DXS7944 X Homo sapiens
 78501 78757 D1S3473 Homo sapiens
 127357 127571 D1S4143 11 Homo sapiens
 148518 148806 SHG-172582 Homo sapiens.

FEATURES
 source location/Qualifiers
 1..161604
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11q"
 /clone="RP11-258013"

BASE COUNT 46275 a 33365 c 33694 g 48270 t
 ORIGIN
 Query Match 70.0%; Score 21; DB 9; Length 161604;
 Best Local Similarity 82.8%; Pred. No. 37;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

HTG; HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens
Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Euteleostomi; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 17876)

Muzny, D.M., Adams, C., Adio-Oduola, A.B., Ali-Osman, F.R., Allen, C.,
Albrooks, S.L., Amaral, H.C., Are, J.R., Banks, T., Barbata, J.,
Benton, J., Bimarge, K., Blankenburg, K., Bonin, D., Bouck, J.,
Bowie, S., Bivete, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthett, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraruto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabis, A., Gao, J., Garcia, A., Garner, T.,
Garcia, N., Gill, R., Gorrell, H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hayes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hoque, M., Hollway, C.,
Hollins, B., Homsl, E., Howard, S., Huber, J., Huliy, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Kovar, C., Kravtsov, J., Kureh, A., Landry, N., Leal, B., Lewis, L.C.,
Lewin, L., Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, W.,
Loulseged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,
Mel, G., Metzger, M., Miner, G., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, S., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwkw, S.,
Oguy, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pulte, M.,
Rui, S., Saverly, G., Scher, S., Scott, G., Shen, H., Shiohara, N.,
Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
Tang, H., Sutton, A., Swatek, A., Tabor, P., Tamerisa, K.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wlaczek, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2. (bases 1 to 17876)
Submitted (07-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 25, 2001 this sequence version replaced gi:12831249.

COMMENT

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: HAMH
Center clone name: RP11-533P2

Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 99% of reads
Assembly program: Phrap; version 0.900329
Consensus quality: 169802 bases at least Q40
Consensus quality: 180431 bases at least Q40
Consensus quality: 185136 bases at least Q20
Estimated insert size: 177736; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 4.2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 42878: contig of 42878 bp in length
* 42879 42978: gap of unknown length
* 42979 62018: contig of 19040 bp in length
* 62019 62119: gap of unknown length
* 62119 80603: contig of 18485 bp in length
* 80604 80704: gap of unknown length
* 80704 98288: contig of 17585 bp in length
* 98289 98389: gap of unknown length
* 98389 113934: contig of 15546 bp in length
* 113935 114034: gap of unknown length
* 114035 125339: contig of 11305 bp in length
* 125340 125439: gap of unknown length
* 125440 138329: contig of 12890 bp in length
* 138330 138429: gap of unknown length
* 138430 145577: contig of 7148 bp in length
* 145578 145678: gap of unknown length
* 145678 150952: contig of 5275 bp in length
* 150953 151052: gap of unknown length
* 151053 155923: contig of 4871 bp in length
* 155924 156023: gap of unknown length
* 156024 159749: contig of 3725 bp in length
* 159750 159848: gap of unknown length
* 159849 163146: contig of 3298 bp in length
* 163147 163246: gap of unknown length
* 163247 166618: contig of 3372 bp in length
* 166619 166718: gap of unknown length
* 166719 170864: contig of 4146 bp in length
* 170865 170964: gap of unknown length
* 170965 173099: contig of 2135 bp in length
* 173100 173199: gap of unknown length
* 173200 175517: contig of 2318 bp in length
* 175518 175617: gap of unknown length
* 175618 17876: contig of 2259 bp in length.

FEATURES

source

1. 17876
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-533P2"
BASE COUNT 45021 a 43543 c 42200 g 45483 t 1629 others
ORIGIN

Query Match 70.0% Score 21; DB 2; Length 17876;
Best Local Similarity 82.8% Pred. No. 37;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGACTCAGTCTTGTCATCTCCTCTCT 30
DB 68919 GGACTCAGTCTTGTCATCTCCTCTCT 68891

RESULT 14

AC130278

LOCUS

DEFINITION

AC130278

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

182695 bp DNA 1linear HTG 09-AUG-2002
unordered pieces.
AC130278.1 GI:22165183
HTG; HTGS_PHASE1; HTGS_DRAFT.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 182695)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP24-393116
Unpublished
2 (bases 1 to 182695)
Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S.,
Barra, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckgeater, B.,
Canarata, J., Chang, J., Chazaro, B., Choepell, Y., Collimore, A.,
Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-pierre, N., Hagos, B.,
Horton, L., Hulme, M., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karats, A., Kells, C., Landers, T., Levine, R., Lindblad-toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Plerie, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Talamas, J.,
Teafaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (09-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Genome Center
Center code: WIR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L25607
Center clone name: 393.1.16

Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 168477 bases at least Q40
Consensus quality: 173980 bases at least Q30
Consensus quality: 176349 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 179195; sum-of-ctrls
Quality coverage: 5.6 in Q20 bases; agarose-fp
Quality coverage: 5.6 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

I 7168: contig of 7168 bp in length
* 7169 7268: gap of 100 bp
* 7269 7773: contig of 505 bp in length
* 7774 7873: gap of 100 bp
* 7874 8549: contig of 676 bp in length
* 8550 8649: gap of 100 bp
* 8650 9264: contig of 615 bp in length
* 9265 9364: gap of 100 bp
* 9365 10058: contig of 694 bp in length
* 10059 10158: gap of 100 bp
* 10159 10902: contig of 744 bp in length
* 10903 11002: gap of 100 bp
* 11003 11256: contig of 254 bp in length
* 11257 11356: gap of 100 bp
* 11357 12233: contig of 877 bp in length

12234 12333: gap of 100 bp
* 12334 13682: contig of 1349 bp in length
* 13683 13782: gap of 100 bp
* 13783 14674: contig of 1092 bp in length
* 14675 14974: gap of 100 bp
* 14975 16211: contig of 1237 bp in length
* 16212 16311: gap of 100 bp
* 16312 17319: contig of 1008 bp in length
* 17320 17419: gap of 100 bp
* 17420 19489: contig of 2070 bp in length
* 19490 19589: gap of 100 bp
* 19590 21253: contig of 1664 bp in length
* 21254 21353: gap of 100 bp
* 21354 22801: contig of 1448 bp in length
* 22802 22901: gap of 100 bp
* 22902 24454: contig of 1533 bp in length
* 24455 24554: gap of 100 bp
* 24555 25711: contig of 1157 bp in length
* 25712 25811: gap of 100 bp
* 25812 28015: contig of 2204 bp in length
* 28016 28115: gap of 100 bp
* 28116 29791: contig of 1676 bp in length
* 29792 29891: gap of 100 bp
* 29892 32114: contig of 2223 bp in length
* 32115 32214: gap of 100 bp
* 32215 32974: contig of 760 bp in length
* 32975 33074: gap of 100 bp
* 33075 33793: contig of 719 bp in length
* 33794 33893: gap of 100 bp
* 33894 37124: contig of 321 bp in length
* 37125 37224: gap of 100 bp
* 37225 40484: contig of 3260 bp in length
* 40485 40584: gap of 100 bp
* 40585 42991: contig of 2407 bp in length
* 42992 43091: gap of 100 bp
* 43092 48293: contig of 5202 bp in length
* 48294 48393: gap of 100 bp
* 48394 53122: contig of 4729 bp in length
* 53123 53222: gap of 100 bp
* 53223 75715: contig of 22493 bp in length
* 75716 75815: gap of 100 bp
* 75816 83666: contig of 7851 bp in length
* 83667 83766: gap of 100 bp
* 83767 92545: contig of 8779 bp in length
* 92546 92645: gap of 100 bp
* 92646 103351: contig of 10706 bp in length
* 103352 103451: gap of 100 bp
* 103452 118198: contig of 14747 bp in length
* 118199 118298: gap of 100 bp
* 118299 134605: contig of 16307 bp in length
* 134606 134705: gap of 100 bp
* 134706 151709: contig of 17004 bp in length
* 151710 151809: gap of 100 bp
* 151810 175524: contig of 23715 bp in length
* 175525 175624: gap of 100 bp
* 175625 182695: contig of 7071 bp in length.

Location/Qualifiers
1. 182695
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-393116"
/clone_lib="RP24-393116 Male Mouse BAC"
1. 7168
/note="assembly-fragment"
vector_side:left"
clone_end:SP6
7269. 7773
/note="assembly-fragment"
7874. 8549
/note="assembly-fragment"
8650. 9264
/note="assembly-fragment"
9365. 10058

FEATURES
source
misc-feature
misc-feature
misc-feature
misc-feature
misc-feature
misc-feature


```

misc_feature /note="assembly_fragment"
10159..10902
misc_feature /note="assembly_fragment"
11003..11256
misc_feature /note="assembly_fragment"
11357..11223
misc_feature /note="assembly_fragment"
12334..13682
misc_feature /note="assembly_fragment"
13783..14874
misc_feature /note="assembly_fragment"
14975..16211
misc_feature /note="assembly_fragment"
16312..17319
misc_feature /note="assembly_fragment"
17420..19489
misc_feature /note="assembly_fragment"
19590..21253
misc_feature /note="assembly_fragment"
21354..22801
misc_feature /note="assembly_fragment"
22902..24454
misc_feature /note="assembly_fragment"
24555..25711
misc_feature /note="assembly_fragment"
25812..28015
misc_feature /note="assembly_fragment"
28116..29791
misc_feature /note="assembly_fragment"
29892..32114
misc_feature /note="assembly_fragment"
32215..332974
misc_feature /note="assembly_fragment"
33075..33793
misc_feature /note="assembly_fragment"
33894..37124
misc_feature /note="assembly_fragment"
37225..40484
misc_feature /note="assembly_fragment"
40385..42991
misc_feature /note="assembly_fragment"
43092..48293
misc_feature /note="assembly_fragment"
48394..53122
misc_feature /note="assembly_fragment"

Query Match 70.0%; Score 21; DB 2; Length 182695;
Best Local Similarity 82.8%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGGACCTAGTCCTGTCATCTACCTTC 29
Db 51988 TGGTCCTCTGCTGTCATCTACCTTC 52016

RESULT 15
AC024545 193700 bp. DNA linear HTG 23-SEP-2000
LOCUS AC024545/c Homo sapiens clone RP11-442P6, WORKING DRAFT SEQUENCE, 39 unordered
DEFINITION Pieces.
AC024545 AC024545.4 GI:10280888
VERSION AC024545.4 GI:10280888
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 193700)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-442P6
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 193700)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

```

TITLE JOURNAL COMMENT

Anderson,S., Baldwin,J., Baran,N., Beda,F., Boguslavsky,L.,
Bouhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Glade,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Klein,J., Lander,T., Larcocque,R., Lehoczy,J., Levine,A.,
Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Mihov,T., Miranda,C., Mlenga,V., Morrow,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
Peterson,K., Plerre,N., Pisan,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,C., Zalnoun,J., Zimmer,A. and
Zody,M.

Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 23, 2000 this sequence version replaced gi:8389572.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L7456
Center clone name: L7456

Summary Statistics
Sequencing vector: MJ3; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 167380 bases at least Q40
Consensus quality: 180250 bases at least Q30
Consensus quality: 185744 bases at least Q20
Insert size: 188000; agarose-ff
Insert size: 189900; sum-of-ctrls
Quality coverage: 3.3 in Q20 bases; agarose-ff
Quality coverage: 3.3 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 1682: contig of 1682 bp in length
* 1683 1782: gap of 100 bp
* 1783 2982: contig of 1200 bp in length
* 2983 3082: gap of 100 bp
* 3083 4543: contig of 1461 bp in length
* 4544 4643: gap of 100 bp
* 4644 5643: contig of 1000 bp in length
* 5644 5743: gap of 100 bp
* 5744 7289: contig of 1546 bp in length
* 7290 7389: gap of 100 bp
* 7390 8911: contig of 1522 bp in length
* 8912 9011: gap of 100 bp
* 9012 11431: contig of 2420 bp in length
* 11432 11531: gap of 100 bp
* 11532 13523: contig of 1992 bp in length
* 13524 13623: gap of 100 bp
* 13624 16363: contig of 2740 bp in length
* 16364 16463: gap of 100 bp

```



```

* 16464 18732: contig of 2269 bp in length
* 18733 18832: gap of 100 bp
* 18833 21778: contig of 2946 bp in length
* 21779 21878: gap of 100 bp
* 21879 25218: contig of 3240 bp in length
* 25219 27553: contig of 2335 bp in length
* 27554 27853: gap of 100 bp
* 27854 30848: contig of 2896 bp in length
* 30850 30949: gap of 100 bp
* 30950 33682: contig of 2733 bp in length
* 33683 33782: gap of 100 bp
* 33783 36641: contig of 2859 bp in length
* 36642 36741: gap of 100 bp
* 36742 40978: contig of 4237 bp in length
* 40979 41078: gap of 100 bp
* 41079 45552: contig of 4474 bp in length
* 45553 45652: gap of 100 bp
* 45653 48341: contig of 2689 bp in length
* 48342 48441: gap of 100 bp
* 48442 53528: contig of 5087 bp in length
* 53529 53628: gap of 100 bp
* 53629 59346: contig of 5718 bp in length
* 59347 59446: gap of 100 bp
* 59447 64005: contig of 4559 bp in length
* 64006 64105: gap of 100 bp
* 64106 68144: contig of 4039 bp in length
* 68145 68244: gap of 100 bp
* 68245 73407: contig of 5163 bp in length
* 73408 73507: gap of 100 bp
* 73508 77763: contig of 4256 bp in length
* 77764 77863: gap of 100 bp
* 77864 82554: contig of 4691 bp in length
* 82555 82654: gap of 100 bp
* 82655 95707: contig of 13053 bp in length
* 95708 95807: gap of 100 bp
* 95808 100939: contig of 5132 bp in length
* 100940 101039: gap of 100 bp
* 101040 106431: contig of 5392 bp in length
* 106432 106531: gap of 100 bp
* 106532 114394: contig of 7863 bp in length
* 114395 114494: gap of 100 bp
* 114495 120736: contig of 6242 bp in length
* 120737 120836: gap of 100 bp
* 120837 127881: contig of 7045 bp in length
* 127882 127981: gap of 100 bp
* 127982 136274: contig of 8293 bp in length
* 136275 136374: gap of 100 bp
* 136375 145186: contig of 8812 bp in length
* 145187 145286: gap of 100 bp
* 145287 153548: contig of 8262 bp in length
* 153549 153648: gap of 100 bp
* 153649 163735: contig of 10107 bp in length
* 163736 163855: gap of 100 bp
* 163856 172952: contig of 9097 bp in length
* 172953 173052: gap of 100 bp
* 173053 187116: contig of 14064 bp in length
* 187117 187216: gap of 100 bp
* 187217 193700: contig of 6484 bp in length.

```

FEATURES

```

source
    1..193700
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="RP11-442P6"
        /clone_lib="RPCT-11 Human Male BAC"
        1..1682
            /note="assembly_fragment"
            clone_end:SP6
            vector_side:left"
        misc_feature
            1783..2982
                /note="assembly_fragment"
                3083..4543
                    /note="assembly_fragment"

```

```

misc_feature      4644..5643
                  /note="assembly_fragment"
misc_feature      5744..7289
                  /note="assembly_fragment"
misc_feature      7390..8911
                  /note="assembly_fragment"
misc_feature      9012..11431
                  /note="assembly_fragment"
misc_feature      11532..13523
                  /note="assembly_fragment"
misc_feature      13624..16363
                  /note="assembly_fragment"
misc_feature      16464..18732
                  /note="assembly_fragment"
misc_feature      18833..21778
                  /note="assembly_fragment"
misc_feature      21879..25118
                  /note="assembly_fragment"
misc_feature      25219..27753
                  /note="assembly_fragment"
misc_feature      27854..30849
                  /note="assembly_fragment"
misc_feature      30950..33682
                  /note="assembly_fragment"
misc_feature      33783..36641
                  /note="assembly_fragment"
misc_feature      36742..40978
                  /note="assembly_fragment"
misc_feature      41079..45552
                  /note="assembly_fragment"
misc_feature      45653..48341
                  /note="assembly_fragment"
misc_feature      48442..53528
                  /note="assembly_fragment"
misc_feature      53629..59346
                  /note="assembly_fragment"
misc_feature      59447..64005
                  /note="assembly_fragment"

```

```

Query Match      70.0%; Score 21; DB 2; Length 193700;
Best Local Similarity 82.8%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY      2 GGACTCAGTCCGTGTCATCCACCTCT 30
DB      67684 GGACTCAGTCTGGTCCCTTCTCT 67656

```

Search completed: July 10, 2003, 19:20:11
Job time : 269.955 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 69.6337 Seconds
(without alignments)
1067.241 Million cell updates/sec

Title: US-10-087-631b-3

Perfect score: 33
Sequence: 1 cgggtacacacgcgtccgacagaccactatgyc 33

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.GeneSeq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	30.4	92.1	36	17	AA139735
C 2	27.4	83.0	29	19	AAV15324
C 3	22	66.7	41	24	AA140116
C 4	22	66.7	68	20	AA560950
C 5	22	66.7	68	20	AA560951
C 6	22	66.7	70	17	AA111268
C 7	22	66.7	120	17	AA109180
C 8	22	66.7	120	18	AA169054
C 9	22	66.7	140	17	AA111269

10	22	66.7	145	17	AA109172	Hepatitis C virus
11	22	66.7	145	24	ABK6834	Hepatitis C virus
C 12	22	66.7	159	14	AA043062	-255 to -62 portio
C 13	22	66.7	159	14	AA043066	-255 to -62 region
C 14	22	66.7	159	14	AA043069	-255 to -62 region
C 15	22	66.7	159	14	AA043071	-255 to -62 region
C 16	22	66.7	163	22	AA510487	HCV 5' UTR domain
C 17	22	66.7	177	15	AA079448	HCV isolate BE82 5
C 18	22	66.7	177	15	AA079449	HCV isolate BE90 5
C 19	22	66.7	177	15	AA079450	HCV isolate BE91 5
C 20	22	66.7	177	15	AA079451	HCV isolate BE92 5
C 21	22	66.7	177	15	AA079452	HCV isolate BE93 5
C 22	22	66.7	177	15	AA079453	HCV isolate BE94 5
C 23	22	66.7	177	15	AA079454	HCV isolate BE98 5
C 24	22	66.7	177	15	AA079455	HCV isolate GB88 5
C 25	22	66.7	177	15	AA079456	HCV isolate GB116
C 26	22	66.7	177	15	AA079457	HCV isolate GB569
C 27	22	66.7	177	15	AA079458	HCV isolate GB558
C 28	22	66.7	177	15	AA079459	HCV isolate CAM600
C 29	22	66.7	177	15	AA079460	HCV isolate CAM736
C 30	22	66.7	177	15	AA068053	HCV isolate GB809
C 31	22	66.7	177	15	AA068057	HCV isolate BE97 5
C 32	22	66.7	177	15	AA068058	HCV isolate GB487
C 33	22	66.7	177	15	AA068059	HCV isolate BE95 5
C 34	22	66.7	177	15	AA068070	HCV isolate BE96 5
C 35	22	66.7	180	13	AA031082	HCV-1 genotype GV
C 36	22	66.7	180	13	AA031083	HCV-1 genotype GV
C 37	22	66.7	184	14	AA043060	-255 to -62 portio
C 38	22	66.7	184	14	AA043061	-255 to -62 portio
C 39	22	66.7	184	14	AA043063	-255 to -62 region
C 40	22	66.7	184	14	AA043064	-255 to -62 region
C 41	22	66.7	184	14	AA043067	-255 to -62 region
C 42	22	66.7	184	14	AA043068	-255 to -62 region
C 43	22	66.7	187	14	AA043069	-255 to -62 portio
C 44	22	66.7	187	14	AA043065	-255 to -62 region
C 45	22	66.7	194	14	AA043058	-255 to -62 portio

ALIGNMENTS

RESULT 1
AA139735/c
ID AA139735 standard; DNA, 36 BP.

AC AA139735;
DT 09-APR-1997 (first entry)
DE Hepatitis C virus PCR probe S1, based on nucleotides 121-157.
XX Hepatitis C virus; HCV, polymerase chain reaction; amplification;
KW replication; non-lymphoblastoid cell; monkey kidney cell;
KM hybridisation probe; ss.
XX Synthetic.
OS
XX
PN WO9624662-A1.
PD 15-AUG-1996.
PE 10-FEB-1995; 95WO-IT00016.
PR 10-FEB-1995; 95WO-IT00016.
PA (CNR) CONSIGLIO NAZ DELLE RICERCHE.
PI Battaglia M, Carloni G, Iacovacci S, Ponzetto A;
PI Ravagnan G;
XX WPI: 1996-384435/38.
XX Replication of hepatitis C virus in non-lymphoblastoid mammalian

PT cells - useful for studies of HCV replication, prodn. of vaccines or
XX viral antigens, etc
XX
PS Disclosure; Page 8; 25pp; English.
XX
CC Hepatitis C virus can be replicated in non-lymphoblastoid mammalian
CC cells by first incubating an HCV sample with the cells until an
CC infecting amount of HCV has been absorbed. Infected cells are then
CC washed and incubated under growth conditions. In an example, the
CC presence of HCV in culture medium of monkey kidney cells inoculated
CC with HCV-infected serum was verified by PCR amplification using two
CC external primers (OU1 and OU2, see AAT39731 and AAT39732) and two
CC internal primers (IN1 and IN2, see AAT39731 and AAT39734). The amplified
CC products were identified by hybridisation to labelled probe S1 (see
CC AAT39735). Small amounts of virus were shown to be released into the
CC culture medium from secondary cultures. Control (non-inoculated) cells
CC were negative for presence of viral genome.
XX
SQ Sequence 36 BP; 9 A; 9 C; 11 G; 7 T; 0 other;
XX
Query Match 92.1%; Score 30.4; DB 17; Length 36;
Best Local Similarity 96.9%; Pred. No. 0.00033;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 CGGTGTTACTCAGCGTTCGACGACCTATG 32
DB 32 CGGTGTTACTCAGCGTTCGACGACCTATG 1
XX
RESULT 2
AAV15324/c
ID AAV15324 standard; DNA; 29 BP.
XX
AC AAV15324;
XX
DT 28-MAY-1998 (first entry)
XX
DE Hepatitis C virus probe HCV40.
XX
KW Hepatitis C virus; HCV; PCR; detection; reverse transcription; probe;
KW enzyme immunoassay; viral RNA; ss.
XX
OS Synthetic.
OS Hepatitis C virus.
XX
PN WC9746716-A1.
XX
PD 11-DEC-1997
XX
PE 03-JUN-1997; 97WO-IT00128.
XX
PR 07-JUN-1996; 96IT-M000404.
XX
PA (WESA) WABCO BV.
XX
PI Bosio P, Clemenza F, Strumia C;
XX
DR WPI; 1998-042222/04.
XX
PT Detection of hepatitis C virus - by reverse transcription,
PT single-step PCR and detection by DNA enzyme immunoassay
XX
PS Disclosure; Page 4; 26pp; English.
XX
CC The present sequence represents a probe involved in the method of
CC the present invention for detecting hepatitis C virus (HCV). The method
CC comprises: (a) reverse-transcribing the viral RNA; (b) amplifying the
CC resulting cDNA by a single polymerase chain reaction in a reaction
CC mixture having a Mg2+/Taq polymerase ratio of about 100 nmole/enzyme
CC unit; and (c) detecting the amplification product by DEIA (DNA enzyme
CC immunoassay) using an oligonucleotide probe. The sensitivity of this
CC method is at least equal to that achievable by more complicated assays
CC using nested PCR.

XX
SQ Sequence 29 BP; 7 A; 7 C; 9 G; 6 T; 0 other;
XX
Query Match 83.0%; Score 27.4; DB 19; Length 29;
Best Local Similarity 96.6%; Pred. No. 0.0072;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 4 TGTTACTACCGTTCGACGACCACTATG 32
DB 29 TGTTACTACCGTTCGACGACCACTATG 1
XX
RESULT 3
AAL40116
ID AAL40116 standard; DNA; 41 BP.
XX
AC AAL40116;
XX
DT 13-SEP-2002 (first entry)
XX
DE Pathogenic microorganism detecting PCR primer SEQ ID No 42.
XX
KW Tuberculosis; Mycobacterium; bovis; BCG; africanum; microti; canottii;
KW monitoring therapy; pathogenic microorganism; PCR; primer; ss.
XX
OS Unidentified.
XX
PN WO200252043-A1.
XX
PD 04-JUL-2002.
XX
PE 26-DEC-2001; 2001WO-JP11422.
XX
PR 26-DEC-2000; 2000JP-0396222.
PR 26-DEC-2000; 2000JP-0396321.
PR 29-JUN-2001; -2001JP-0199552.
PR 13-SEP-2001; 2001JP-0278920.
XX
PA (TAKI) TAKARA SHUZO CO LTD.
XX
PI Shinada M, Hino F, Kato I;
XX
DR WPI; 2002-500769/53.
XX
PT Detecting pathogenic microorganisms with oligonucleotide probes and
PT primers, useful in disease diagnosis and monitoring therapy
XX
PS Claim 57; Page 97; 106pp; Japanese.
XX
CC The invention relates to a probe containing a 410 or 20 base pair
CC sequence, given in the specification. It is capable of detecting the
CC tuberculosis bacterial group including Mycobacterium tuberculosis,
CC Mycobacterium bovis BCG, Mycobacterium africanum, Mycobacterium microti
CC and/or Mycobacterium canottii. The method is useful in disease diagnosis
CC and monitoring therapy. This polynucleotide sequence represents a PCR
CC primer relating to the detection of pathogenic microorganisms of the
CC invention.
XX
SQ Sequence 41 BP; 7 A; 14 C; 9 G; 11 T; 0 other;
XX
Query Match 66.7%; Score 22; DB 24; Length 41;
Best Local Similarity 97.1%; Pred. No. 2;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
OY 1 CGGTGTTACTCAGC-GTTCGACGACCACTATG 33
DB 5 CGGTGTTACTCAGCGTTCGACGACCACTATG 38
XX
RESULT 4
AAX60950/c
ID AAX60950 standard; DNA; 68 BP.
XX

AC AAX60950;
XX
DT 16-AUG-1999 (first entry)
XX
DE Hepatitis C virus (HCV) biotinylated probe 3CH.
XX
KW Nucleic acid detection; pathogen; bacteria; virus; hepatitis C virus;
KW HCV; hepatitis B; HBV; hepatitis G; HGV; HIV; fungus; protozoa; ss;
KW parasite; mycoplasma; genetic mutation; food contamination; probe.
XX
OS Synthetic.
OS Hepatitis C virus.
PN W09928503-A1.
PN
PD 10-JUN-1999.
PD
PF 16-NOV-1998; 98WO-US24494.
XX
PR 03-DEC-1997; 97IT-RM00749.
XX
PA (DIAS-) DIASORIN INT INC.
XX
PI Mantero G, Priml D;
XX
DR WPI; 1999-371139/31.
XX
PT Detection of single-stranded polynucleotide analytes
PS
PS Claim 6; Page 65; 73pp; English.
CC The invention relates to a new method for detection of single-stranded
CC (ss) polynucleotide analytes that comprises using ss polynucleotide
CC probes which hybridize to the analyte and are bound to a solid support
CC where double-stranded (ds) polynucleotides are detected. The method can
CC be used for detecting ss PN analytes for the detection of pathogens such
CC as bacteria, viruses such as hepatitis C (HCV), hepatitis B (HBV),
CC hepatitis G (HGV), or HIV, fungi, protozoa, parasites or mycoplasma. The
CC method can be used to detect the presence of genetic mutations which have
CC diagnostic or prognostic value. Pathogen contamination of food and drink
CC supplies can also be detected using the method. The method provides for
CC the sensitive and specific detection of ss PN analytes at concentration
CC as low as 0.1fg.
XX
SQ Sequence 68 BP; 14 A; 17 C; 24 G; 13 T; 0 other;
XX
Query Match 66.7%; Score 22; DB 20; Length 68;
Best Local Similarity 97.1%; Pred. No. 2.2;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
OY 1 CGGTGTAACACG-GTTCCGACGACCACTATGCG 33
|||||
DB 40 CGGTGTAACACCGGTTCCGACGACCACTATGCG 7
|||||
RESULT 5
ID AAX60951 standard; DNA; 68 BP.
XX
AC AAX60951;
XX
DT 16-AUG-1999 (first entry)
XX
DE Hepatitis C virus (HCV) biotinylated probe alpha-3CH.
XX
KW Nucleic acid detection; pathogen; bacteria; virus; hepatitis C virus;
KW HCV; hepatitis B; HBV; hepatitis G; HGV; HIV; fungus; protozoa; ss;
KW parasite; mycoplasma; genetic mutation; food contamination; probe.
XX
OS Synthetic.
OS Hepatitis C virus.
PN W09928503-A1
PN

[illegible]

XX The present sequence is a specifically claimed example of RNA that
 CC is complementary (i.e. antisense) to part of the 5' untranslated
 CC region of the hepatitis C virus genome sequence contained in clone
 CC 2-1. The 5'-UTR includes several stem-loop sequences. The antisense
 CC RNA is useful for inhibiting expression of HCV structural genes and
 CC thereby inhibiting viral replication in vivo. The antisense therapy
 CC can be used in addition to conventional interferon treatment of HCV
 CC infections.

XX
 SQ Sequence 70 BP; 10 A; 21 C 25 G; 14 U; 0 other;

Query Match
 Best Local Similarity 66.7%; Score 22; DB 17; Length 70;
 Matches 26; Conservative 7; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGACTCACC-GTTCGCGAGACCACTATGCG 33
 10 CGGUGACUCACCGGUCGCGAGACCACTATGCG 43

RESULT 7

ID AAT09180 standard; DNA; 120 BP.

AC AAT09180;

DE 14-AUG-1996 (first entry).

XX Hepatitis C virus specific ligated amplified sequence.

KW Ligase dependent polymerase chain reaction; LD-PCR; probe; hybridisation;
 KM ligand binding pair; ligase; paramagnetic bead; primer; amplification;
 XX hepatitis; untranslated region; UTR; RNA; ss.

OS Synthetic.

PN WO9535390-A1.

PD 28-DEC-1995.

PF 14-JUN-1995; 95WO-US07671.

PR 22-JUN-1994; 94US-0263937.

PA (MOUN) MOUNT SINAI SCHOOL MEDICINE.

PI Zhang DY;

DR WPI; 1996-058427/06.

XX

XX

XX

XX

XX

XX

XX

XX

CC

Example 5; Page 54; 100pp; English.

A novel method of detecting a target nucleic acid (TNA) sequence
 involves use of the ligase dependent polymerase chain reaction method
 (LD-PCR). In this method, two probes are provided. The first probe
 contains a region at the 5' end which is complementary and will hybridise
 with the TNA, the 3' end of the first probe is generic and is bound to
 one half of a ligand binding pair (LBP). The second probe contains a
 region at the 5' end which is complementary to a region in the TNA which
 is immediately adjacent to the complementary region of the first probe.
 When the probes are bound to the TNA, they can be ligated together using
 a conventional ligase. The TNA:ligated probe complex can be isolated by
 binding the first probe to a paramagnetic bead to which is attached the
 second half of the LBP. The TNA can be dissociated from the ligated
 probe complex which can then be detected either by a label attached to
 the second probe, by using an external probe or by PCR using the ligated
 probe as a template.

The capture probes AAT09176-7 are used to isolated a region of the

CC Hepatitis C virus 5' untranslated region. This is the sequence of
 CC the ligated amplified region created by the ligation of the two
 CC amplification probes amp-probe-2 and -2A (AAT09178-9). The ligated
 CC sequence can subsequently be detected by PCR amplification with
 CC the primers AAT09181-3.

XX
 SQ Sequence 120 BP; 26 A; 34 C; 29 G; 31 T; 0 other;

Query Match
 Best Local Similarity 66.7%; Score 22; DB 17; Length 120;
 Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGACTCACC-GTTCGCGAGACCACTATGCG 33
 64 CGGTGACTCACC-GTTCGCGAGACCACTATGCG 97

RESULT 8

ID AAT69054/c standard; DNA; 120 BP.

AC AAT69054;

DE 16-JUL-1997 (first entry)

XX Hepatitis C virus fragment.

KW Polymerase chain reaction; amplification; hepatitis B virus; HBV;
 KM hepatitis C virus; HCV; simultaneous analysis; nested PCR; ss.

OS Synthetic.

PN KR9508291-B1.

PD 27-JUL-1995.

PF 22-APR-1992; 92KR-0006826.

PR 22-APR-1992; 92KR-0006826.

PA (MOUN/) MUN I.

PI Mun I;

DR WPI; 1997-191071/17.

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC

Example 3; Page 7; 12pp; Korean.

Hepatitis B virus (HBV) and hepatitis C virus (HCV) can be analysed
 simultaneously using specific sets of PCR primers. Specifically,
 viral nucleic acids are isolated by treating human serum with a
 guanidine isothiocyanate containing solution. HBV and HCV nucleic
 acids are then absorbed on silica and a first polymerase chain reaction
 (PCR) is performed using primers HB-1 and HB-2 (AAT65089 and AAT65090) and
 CC HC-1 and HC-2 (AAT65093 and AAT65094). A second PCR is then performed
 CC with the primers HB-3 and HB-4 (AAT65091 and AAT65092) and HC-3 and HC-4
 CC (AAT65095 and AAT65096). A PCR fragment of 258 bp is produced for HBV and
 CC a PCR fragment of 163 bp is produced for HCV. The present sequence
 XX represents a fragment from HCV.

SQ Sequence 120 BP; 21 A; 37 C; 34 G; 22 T; 6 other;

Query Match
 Best Local Similarity 66.7%; Score 22; DB 18; Length 120;
 Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGACTCACC-GTTCGCGAGACCACTATGCG 33
 58 CGGTGACTCACC-GTTCGCGAGACCACTATGCG 25

PA (BRAN)/ BRANDWEIN M.
 XX (HSUI)/ HSUI T C H.
 PI Zhang DY, Brandwein M, Hsuih TCH;
 XX WPI; 2002-508508/54.
 DR
 XX
 PT Detecting a target nucleic acid by a ramification-extension
 PT amplification method, involves carrying out the steps of capturing,
 PT carrying out ligand-dependent amplification and detection of the
 XX target nucleic acid
 XX
 PS Example 4; Page 78; 160pp; English.

CC This invention relates to a novel method for detecting target nucleic
 CC acid using a Capture/Amplification (C/A) probe and an amplification
 CC probe (A-probe), where C/A probe contains generic sequences and
 CC sequences complementary to target nucleic acid sequence, and also
 CC contains a ligand bound to the non-complementary sequence of probe; the
 CC ligand is capable of forming an affinity pair with a ligand binding
 CC moiety coated on matrix. The method of the invention may be used for
 CC detecting a target nucleic acid in sample. Preferably, the method is
 CC useful for detecting and quantitating nucleic acids from pathogenic
 CC microorganisms, from samples from patients with infectious diseases.
 CC The method allows for rapid, sensitive and standardised detection and
 CC quantification of nucleic acids from pathogenic microorganisms from
 CC samples from patients with infectious diseases. The method simplifies
 CC the target nucleic acid isolation procedure, which can be performed in
 CC microtubes, microchips or micro-well plates, if desired. The method
 CC also allows the direct detection of RNA by probe amplification without
 CC the need for DNA template production. The assembly of the amplifiable
 CC DNA by ligation increases specificity, and makes possible the detection
 CC of a single mutation in a target. The method can be easily automated.
 CC The present sequence represents the nucleic acid detection method
 CC hepatitis C virus ligated amplification sequence used in an example of
 CC the method of the invention
 XX
 SO Sequence 145 BP; 29 A; 40 C; 40 G; 36 T; 0 other;

Query Match. 66.7%; Score 22; DB 24; Length 145;
 Best Local Similarity 97.1%; Pred. No. 2.4;
 Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTCACC-GTTCGCCAGACCACTATGCG 33
 DB 72 CGGTGTACTCACC-GTTCGCCAGACCACTATGCG 105

RESULT 12
 AAQ43062/C

ID AAQ43062 standard; cDNA; 159 BP.

AC AAQ43062;

DT 23-SEP-1993 (first entry)

DE -255 to -62 portion of 5' non-coding region of HCV K2a.

KW Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
 KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3; ss.

OS Hepatitis C virus.

PN WO9310239-A.

PD 27-MAY-1993.

PF 20-NOV-1992; 92MO-GB02143.

PR 21-NOV-1991; 91GB-0024696.

PR 24-JUN-1992; 92GB-0013362.
 (COMM-) COMMON SERVICES AGENCY.

XX Chan S, Simmonds P, Yap PL;
 PI WPI; 1993-182554/22.
 XX
 DR
 XX

PT DNA encoding antigenic peptide(s) of new types of hepatitis C
 PT virus - for diagnosing and treating HCV infection, screening
 PT blood samples and identifying different HCV types
 XX
 PS Disclosure; Fig 1; 120pp; English.

CC The sequences given in AAQ43058-75 show the -255 to -62 non-coding
 CC region of hepatitis C virus (HCV) samples from 18 blood donors and
 CC other HCV variants. Analysis of this region revealed the existence
 CC of three distinct groups of HCV differing by 9-14% in nucleotide
 CC sequence. Two of the groups identified were similar to those of HCV
 CC variants termed type 1 and 2, whilst the third appeared to represent
 CC a novel virus type. Comparison of other regions of the genome, eg.
 CC the NS-5 region (see also AAR3/923-26), showed a high degree of sequence
 CC diversity with type 3 being phylogenetically different to type 1 and 2.
 CC The same degree of differentiation was noted in the NS-3 (see AAR3/927-30)
 CC and core region (see AAR3/931) between type 3 and type 1 sequences.
 XX
 SO Sequence 159 BP; 29 A; 49 C; 48 G; 33 T; 0 other;

Query Match. 66.7%; Score 22; DB 14; Length 159;
 Best Local Similarity 97.1%; Pred. No. 2.5;
 Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTCACC-GTTCGCCAGACCACTATGCG 33
 DB 84 CGGTGTACTCACC-GTTCGCCAGACCACTATGCG 51

RESULT 13
 AAQ43066/C

ID AAQ43066 standard; cDNA; 159 BP.

AC AAQ43066;

DT 23-SEP-1993 (first entry)

DE -255 to -62 region of 5' non-coding region of HCV K2b-1.

KW Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
 KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3; ss.

OS Hepatitis C virus.

PN WO9310239-A.

PD 27-MAY-1993.

PF 20-NOV-1992; 92MO-GB02143.

PR 21-NOV-1991; 91GB-0024696.

PR 24-JUN-1992; 92GB-0013362.
 (COMM-) COMMON SERVICES AGENCY.

PI Chan S, Simmonds P, Yap PL;

PI WPI; 1993-182554/22.

PT DNA encoding antigenic peptide(s) of new types of hepatitis C
 PT virus - for diagnosing and treating HCV infection, screening
 PT blood samples and identifying different HCV types
 XX
 PS Disclosure; Fig 1; 120pp; English.

CC The sequences given in AAQ43058-75 show the -255 to -62 non-coding
 CC region of hepatitis C virus (HCV) samples from 18 blood donors and
 CC other HCV variants. Analysis of this region revealed the existence

CC of three distinct groups of HCV differing by 9-14% in nucleotide
 CC sequence. Two of the groups identified were similar to those of HCV
 CC variants termed type 1 and 2, whilst the third appeared to represent
 CC a novel virus type. Comparison of other regions of the genome, eg.
 CC the NS-5 region (see also AAR37923-26), showed a high degree of sequence
 CC diversity with type 3 being phylogenetically different to type 1 and 2.
 CC The same degree of differentiation was noted in the NS-3 (see AAR37927-30)
 CC and core region (see AAR37931) between type 3 and type 1 sequences.
 XX
 SO Sequence 159 BP; 30 A; 48 C; 47 G; 34 T; 0 other;

Query Match 66.7%; Score 22; DB 14; Length 159;
 Best Local Similarity 97.1%; Pred. No. 2.5;
 Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 33
 DB 84 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 51

RESULT 14
 AAQ43069/C
 ID AAQ43069 standard; CDNA: 159 BP.

XX AAQ43069;
 AC
 XX 23-SEP-1993 (first entry)

DE -255 to -62 region of 5' non-coding region of HCV Pt-1.

XX Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
 KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3; ss.

XX Hepatitis C virus.

XX W09310239-A.

XX 27-MAY-1993.

XX 20-NOV-1992; 92WO-GB02143.

XX 21-NOV-1991; 91GB-0024696.

XX 24-JUN-1992; 92GB-0013362.

XX (COMM-) COMMON SERVICES AGENCY.

XX Chan S, Simmonds P, Yap PL;

XX WPI; 1993-182554/22.

XX DNA encoding antigenic peptide(s) of new types of hepatitis C
 PT virus - for diagnosing and treating HCV infection, screening
 PT blood samples and identifying different HCV types

XX Disclosure: Fig 1; 120pp; English.

XX The sequences given in AAQ43058-75 show the -255 to -62 non-coding
 CC region of hepatitis C virus (HCV) samples from 18 blood donors and
 CC other HCV variants. Analysis of this region revealed the existence
 CC of three distinct groups of HCV differing by 9-14% in nucleotide
 CC sequence. Two of the groups identified were similar to those of HCV
 CC variants termed type 1 and 2, whilst the third appeared to represent
 CC a novel virus type. Comparison of other regions of the genome, eg.
 CC the NS-5 region (see also AAR37923-26), showed a high degree of sequence
 CC diversity with type 3 being phylogenetically different to type 1 and 2.
 CC The same degree of differentiation was noted in the NS-3 (see AAR37927-30)
 CC and core region (see AAR37931) between type 3 and type 1 sequences.

XX Sequence 159 BP; 30 A; 48 C; 50 G; 31 T; 0 other;

Query Match 66.7%; Score 22; DB 14; Length 159;
 Best Local Similarity 97.1%; Pred. No. 2.5;
 Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 33
 DB 84 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 51

RESULT 15
 AAQ43071/C
 ID AAQ43071 standard; CDNA: 159 BP.

XX AAQ43071;

XX 23-SEP-1993 (first entry)

DE -255 to -62 region of 5' non-coding region of HCV J1.

XX Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
 KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3; ss.

XX Hepatitis C virus.

XX W09310239-A.

XX 27-MAY-1993.

XX 20-NOV-1992; 92WO-GB02143.

XX 21-NOV-1991; 91GB-0024696.

XX 24-JUN-1992; 92GB-0013362.

XX (COMM-) COMMON SERVICES AGENCY.

XX Chan S, Simmonds P, Yap PL;

XX WPI; 1993-182554/22.

XX DNA encoding antigenic peptide(s) of new types of hepatitis C
 PT virus - for diagnosing and treating HCV infection, screening
 PT blood samples and identifying different HCV types

XX Disclosure: Fig 1; 120pp; English.

XX The sequences given in AAQ43058-75 show the -255 to -62 non-coding
 CC region of hepatitis C virus (HCV) samples from 18 blood donors and
 CC other HCV variants. Analysis of this region revealed the existence
 CC of three distinct groups of HCV differing by 9-14% in nucleotide
 CC sequence. Two of the groups identified were similar to those of HCV
 CC variants termed type 1 and 2, whilst the third appeared to represent
 CC a novel virus type. Comparison of other regions of the genome, eg.
 CC the NS-5 region (see also AAR37923-26), showed a high degree of sequence
 CC diversity with type 3 being phylogenetically different to type 1 and 2.
 CC The same degree of differentiation was noted in the NS-3 (see AAR37927-30)
 CC and core region (see AAR37931) between type 3 and type 1 sequences.

XX Sequence 159 BP; 30 A; 49 C; 50 G; 30 T; 0 other;

Query Match 66.7%; Score 22; DB 14; Length 159;
 Best Local Similarity 97.1%; Pred. No. 2.5;
 Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 33
 DB 84 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 51

Search completed: July 10, 2003, 19:52:33
 Job time : 70.6337 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 508.537 Seconds
(without alignments)
1067.241 Million cell updates/sec

Title: US-10-087-631B-10
Perfect score: 241
Sequence: 1 gcagaagcgtctagccatg.....ctgcctgataggtgtctgc 241

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: N_Geneseq_101002.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	241	100.0	305	24	ABN79970
2	241	100.0	328	24	ABLA6275
3	241	100.0	341	20	AAZ08979
4	241	100.0	341	20	AAZ24834
5	241	100.0	341	20	AAZ24835
6	241	100.0	342	21	AAZ57773
7	241	100.0	393	24	ABA56260
8	241	100.0	412	24	ABA56259
9	241	100.0	483	13	AAO32444

10	241	100.0	483	13	AAO32445	HCV core-envelope
11	241	100.0	587	21	AAZ57395	Hepatitis C virus
12	241	100.0	686	15	AAO44921	Hepatitis C virus
13	241	100.0	703	21	AAZ57396	Hepatitis C virus
14	241	100.0	713	21	AAO8097	Hepatitis type C v
15	241	100.0	720	21	ABLS2861	Hepatitis C virus
16	241	100.0	780	21	AAZ57789	Hepatitis C virus
17	241	100.0	923	17	AAZ28348	Hepatitis C virus
18	241	100.0	1554	13	AAO32448	HCV core-envelope
19	241	100.0	1880	13	AAO24467	NANB hepatitis vir
20	241	100.0	2033	15	AAO64913	Hepatitis C virus
21	241	100.0	2033	16	AAO86788	Hepatitis C virus
22	241	100.0	2540	13	AAO29628	Hepatitis C virus
23	241	100.0	2540	14	AAO43889	NANB hepatitis vir
24	241	100.0	2540	15	AAO63753	NANBH genomic fra
25	241	100.0	3401	15	AAO64069	Non A, non-B hepat
26	241	100.0	3401	16	AAZ0387	5'UTR/CORE/ENV/NS1
27	241	100.0	4987	15	AAO65322	Vaccinia virus pro
28	241	100.0	4987	16	AAO86799	Vaccinia virus vec
29	241	100.0	7141	24	AAO25333	Hepatitis C virus
30	241	100.0	7789	24	AAO25330	Hepatitis C virus
31	241	100.0	7848	24	AAO25323	Hepatitis C virus
32	241	100.0	7987	24	AAO25321	Hepatitis C virus
33	241	100.0	7987	24	AAO25324	Hepatitis C virus
34	241	100.0	7987	24	AAO25329	Hepatitis C virus
35	241	100.0	7989	21	AAO89668	Hepatitis C virus
36	241	100.0	7989	24	AAO25322	Hepatitis C virus
37	241	100.0	7989	24	AAO25325	Hepatitis C virus
38	241	100.0	7989	24	AAO25326	Hepatitis C virus
39	241	100.0	7991	24	AAO47279	Hepatitis C virus
40	241	100.0	7992	24	AAO47276	Hepatitis C virus
41	241	100.0	7992	24	AAO47277	Hepatitis C virus
42	241	100.0	7992	24	AAO47280	Hepatitis C virus
43	241	100.0	7992	24	AAO47281	Hepatitis C virus
44	241	100.0	7995	24	AAO47278	Hepatitis C virus
45	241	100.0	8001	21	AAO89667	Hepatitis C virus

ALIGNMENTS

RESULT 1	ABN79970	standard; DNA: 305 BP.
ID	ABN79970	
AC	ABN79970	
DT	15-JUL-2002	(first entry)
XX		
XX		
DE	Hepatitis C virus 5' untranslated region genotype 1b.	
XX		
KW	Single nucleotide polymorphism; nucleic acid typing; hepatitis C virus;	
KW	tissue typing; untranslated region; UTR; ds; HCV.	
XX		
OS	Hepatitis C virus.	
XX		
PN	WO200220837-A2.	
XX		
PD	14-MAR-2002.	
XX		
PF	10-SEP-2001; 2001WO-GB04042.	
XX		
PR	08-SEP-2000; 2000GB-0022069.	
XX		
PA	(PYRO-) PYROSEQUENCING AB.	
PA	(STRD) UNIV LELAND STANFORD JUNIOR.	
PA	(GARD) GARDNER R.	
XX		
PI	Romagnh M, Ekstroem B, Pourmand N;	
XX		
DR	WPI: 2002-393849/42.	
XX		
PT	Typing nucleic acid for obtaining information about several variable	

PT sites involves simultaneously or sequentially performing two or more
PT primer extension reactions, and determining the pattern of nucleotide
incorporation.

Example 1; Fig 2; 86pp; English.

CC The invention relates to a novel method for obtaining typing information
CC about several variable sites within target nucleic acid, or typing one or
CC more nucleic acid molecules. The methods of the invention are useful for
CC typing one or more nucleic acid molecules containing two or more variable
CC sites, preferably nucleic acid molecules containing three or more
CC variable sites are typed where three or more primer extension reactions
CC are performed. The method is also useful for diagnosis of pathological
CC conditions characterized by the presence of specific nucleic acid
CC molecules. The methods are particularly suited for identifying
CC of polymorphisms, tissue typing or in clinical applications. The sequence
CC represents the 5' untranslated region (UTR) of a hepatitis C virus (HCV)
CC genotype, amplified in the invention to type HCV-positive sera.

SQ Sequence 305 BP; 56 A; 87 C; 98 G; 64 T; 0 other;

Query Match 100.0%; Score 241; DB 24; Length 305;
Best Local Similarity 100.0%; Pred. No. 2.9e-62;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGGCTAGCCTAGGCTTGTAGTGTCTGAGCTTCCAGACCCCT 60
DB 26 GCAGAAAGGCTAGCCTAGGCTTGTAGTGTCTGAGCTTCCAGACCCCT 85
QY 61 CCCGGAGAGCCATAGTGTCTGCGAACCCTGAGTACCGGAATTGCCAGACACC 120
DB 86 CCCGGAGAGCCATAGTGTCTGCGAACCCTGAGTACCGGAATTGCCAGACACC 145
QY 121 GGGTCTCTTCTGTGATCAACCCGCTCAATGCTGAGATTGGCGTCCCGCGAGAC 180
DB 146 GGGTCTCTTCTGTGATCAACCCGCTCAATGCTGAGATTGGCGTCCCGCGAGAC 205
QY 181 TCGTAGCCGAGTAGTGTGGTTCGCGAAGGCTTGTGATCTGCTGATAGGCTGTTG 240
DB 206 TCGTAGCCGAGTAGTGTGGTTCGCGAAGGCTTGTGATCTGCTGATAGGCTGTTG 265
QY 241 C 241
DB 266 C 266

RESULT 2
AB146275 standard; RNA; 326 BP.

XX ID AB146275
XX AC AB146275;
XX DT 26-APR-2002 (first entry)
XX DE Hepatitis C virus subtype 1a RNA sequence SEQ ID NO:242.
XX KW Nucleic acid accessible hybridisation site; detection; hybridisation;
XX gene; ss.
XX OS Hepatitis C virus.
XX PN WO200198537-A2.
XX PD 27-DEC-2001.
XX PF 15-JUN-2001; 2001WO-US19401.
XX PR 17-JUN-2000; 2000US-212308P.
XX PR 15-JUN-2001; 2001US-0212308.
XX PA (THIR-) THIRD WAVE TECHNOLOGIES INC.

XX LYamichev V, Allawi H, Dong F, Neil BP, Vener IT;
XX WPI; 2002-049698/06.

PT Identifying oligonucleotides hybridizing to nucleic acids containing
PT secondary structure, useful in clinical diagnosis, comprises
PT identifying primers that interact with the target to form an extension
PT product under amplification conditions.

PS Claim 48; Fig 74; 409pp; English.

CC The present invention describes a method for identifying oligonucleotides
CC with desired hybridisation properties to nucleic acid targets containing
CC secondary structure. The method comprises amplifying a target nucleic
CC acid having at least one accessible and one inaccessible site. Primers
CC that form an extension product are identified as the oligonucleotides
CC which can interact with the folded target nucleic acid. Oligonucleotides
CC from the present invention can be used in novel detection methods for
CC clinical diagnostic purposes, including the detection and identification
CC of pathogenic organisms (e.g. HIV). The method allows the ability to
CC rapidly analyse nucleic acid structures. AB146034 to AB146367 represent
CC sequences used in the exemplification of the present invention.

SQ Sequence 328 BP; 63 A; 95 C; 102 G; 68 U; 0 other;

Query Match 100.0%; Score 241; DB 24; Length 328;
Best Local Similarity 78.8%; Pred. No. 3e-62;
Matches 190; Conservative 51; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGGCTAGCCTAGGCTTGTAGTGTCTGAGCTTCCAGACCCCT 60.
DB 50 GCAGAAAGGCTAGCCTAGGCTTGTAGTGTCTGAGCTTCCAGACCCCT 109
QY 61 CCCGGAGAGCCATAGTGTCTGCGAACCCTGAGTACCGGAATTGCCAGACACC 120
DB 110 CCCGGAGAGCCATAGTGTCTGCGAACCCTGAGTACCGGAATTGCCAGACACC 169
QY 121 GGGTCTCTTCTGTGATCAACCCGCTCAATGCTGAGATTGGCGTCCCGCGAGAC 180
DB 170 GGGTCTCTTCTGTGATCAACCCGCTCAATGCTGAGATTGGCGTCCCGCGAGAC 229
QY 181 TCGTAGCCGAGTAGTGTGGTTCGCGAAGGCTTGTGATCTGCTGATAGGCTGTTG 240
DB 230 TCGTAGCCGAGTAGTGTGGTTCGCGAAGGCTTGTGATCTGCTGATAGGCTGTTG 289
QY 241 C 241
DB 290 C 290

RESULT 3
AAZ08979 standard; DNA; 341 BP.

XX ID AAZ08979
XX AC AAZ08979;
XX DT 02-NOV-1999 (first entry)
XX DE Hepatitis C virus 5' UTR used as promoter for a HCV genetic vaccine.
XX KW Genetic vaccine; immunisation; humoral response; cellular response;
XX non-structural protein; NS protein; antigen; ds.
XX OS Hepatitis C virus.
XX PN WO9938880-A1.
XX PD 05-AUG-1999.
XX PF 28-JAN-1999; 99WO-US01823.
XX PR 30-JAN-1998; 98US-0073156.

XX (GEMO) GEN HOSPITAL CORP.
PA Encke J, Mands J;
PI WPI: 1999-494077/41.

XX New hepatitis virus nucleic acids for, e.g. inducing an immune
PT response against the virus
PS Claim 6; Page 10-11: 41pp; English.

XX This sequence represents the hepatitis C virus (HCV) 5' untranslated
CC region (5' UTR). This is preferred for use as a promoter in a
CC recombinant DNA vaccine against hepatitis C. The genetic
CC vaccine comprises regulatory elements capable of directing
CC expression in humans and one or more HCV non-structural
CC (NS) proteins. The HCV genome encodes three NS proteins: NS3, a serine
CC protease/viral helicase; NS4; and NS5, the viral RNA-dependent RNA
CC polymerase. In HCV infection, these and other virally encoded proteins
CC are produced by proteolytic processing of a precursor polypeptide.
CC However, for use in a genetic vaccine, the sequences coding for NS
CC proteins had to be engineered so that each NS sequence had its own
CC initiation and stop codons. Restriction sites were also engineered
CC into the gene fragments to aid subcloning. The gene fragments were
CC isolated and mutated using PCR. The NS3 coding sequence was modified
CC using PCR primers AA208980 and AA208981, the NS4 coding sequence
CC modified using AA208984 and AA208985, and the NS5 coding sequence
CC modified using AA208986 and AA208987. Studies in mice immunised with
CC constructs expressing one NS protein were found to elicit strong
CC antigen-specific immune responses in both arms of the immune system. This
CC demonstrated that the NS proteins are better antigens for stimulating
CC humoral immune responses as compared with previous studies using the HCV
CC core structural protein. The genetic vaccine is useful for inducing
CC an immune response (cellular or humoral) against hepatitis C virus in
CC a human uninfected by the virus, and for immunising a human susceptible
CC to hepatitis C viral infection by inducing an immune response. The
CC composition is also useful for treating a human infected with hepatitis
CC C virus, by induction of an immune response. The advantage of this
CC method of immunisation compared with immunisations with soluble
CC recombinant proteins or peptides, is the ability to induce a strong
CC inflammatory CD4+ T cell response as well as cytotoxic T cell activity.
CC In addition, the new recombinant genetic vaccine is more suitable for
CC immunisation, unlike synthetic peptides which only have a limited
CC number of epitopes available for stimulation of the host response.

XX Sequence 341 BP; 63 A; 103 C; 106 G; 69 T; 0 other;

Query Match 100.0%; Score 241; DB:20; Length 341;
Best Local Similarity 100.0%; Pred. No. 3e-62;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGCTTATGATGATGCTGTCGAGCCCTCCAGAGACCCCCCT 60
DB 68 GCAGAAAGCGTCTAGCCATGCGCTTATGATGATGCTGTCGAGCCCTCCAGAGACCCCCCT 127
QY 61 CCCGGAGAGCCATAGTGTGCGGAAACCGGTAGTACACCGGAATTGGCAGACGACC 120
DB 128 CCCGGAGAGCCATAGTGTGCGGAAACCGGTAGTACACCGGAATTGGCAGACGACC 187
QY 121 GGGTCCTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCGTGCACCCCGAGAC 180
DB 188 GGGTCCTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCGTGCACCCCGAGAC 247
QY 181 TGTAGCCGAGTACTGTTGGTGCAGAAAGCGCTTGTGATGCTCCTGATAGGGTGTG 240
DB 248 TGTAGCCGAGTACTGTTGGTGCAGAAAGCGCTTGTGATGCTCCTGATAGGGTGTG 307
QY 241 C 241
DB 308 C 308

RESULT 4
AA24834
ID AA24834 standard; DNA; 341 BP.

XX AC AA24834;
XX 21-JUN-1999 (first entry)

DE Infectious hepatitis C virus genotype 1a/1b 5'UTR consensus.

XX HCV; infectious clone; infection; diagnosis; therapy; vaccine;
KW screening; assay; antiviral; virucide; ss.

OS Hepatitis C virus.

XX W09904008-A2.

XX 28-JAN-1999.

XX 16-JUL-1998; 98WO-0514688.

XX 27-JAN-1998; 98US-0014416.

XX 18-JUL-1997; 97US-0053062.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Buhk J, Emerson SU, Purcell RH, Yanagi M;
PI WPI: 1999-132252/11.

XX New isolated hepatitis C virus nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of HCV
PT infections and for developing screening assays

XX Example 7; Fig 11; 126pp; English.

XX The present sequence comprises a consensus of the 5' untranslated
CC regions (5'UTR) of hepatitis C virus (HCV) infectious clones of
CC genotype 1a (pcv-H77C, see AA24836) and 1b (pcv-J4L65, see AA24835).
CC The invention discloses nucleic acid sequences (see AA24832-33 and
CC AA24843) which encode infectious HCV viruses, and the use of these
CC sequences, and polypeptides (see AA248020-22) encoded by them, in the
CC development of vaccines and diagnostics for HCV and in the
CC development of screening assays for the identification of antiviral
CC agents for HCV.

XX Sequence 341 BP; 63 A; 103 C; 106 G; 69 T; 0 other;

Query Match 100.0%; Score 241; DB:20; Length 341;
Best Local Similarity 100.0%; Pred. No. 3e-62;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGCTTATGATGATGCTGTCGAGCCCTCCAGAGACCCCCCT 60
DB 68 GCAGAAAGCGTCTAGCCATGCGCTTATGATGATGCTGTCGAGCCCTCCAGAGACCCCCCT 127
QY 61 CCCGGAGAGCCATAGTGTGCGGAAACCGGTAGTACACCGGAATTGGCAGACGACC 120
DB 128 CCCGGAGAGCCATAGTGTGCGGAAACCGGTAGTACACCGGAATTGGCAGACGACC 187
QY 121 GGGTCCTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCGTGCACCCCGAGAC 180
DB 188 GGGTCCTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCGTGCACCCCGAGAC 247
QY 181 TGTAGCCGAGTACTGTTGGTGCAGAAAGCGCTTGTGATGCTCCTGATAGGGTGTG 240
DB 248 TGTAGCCGAGTACTGTTGGTGCAGAAAGCGCTTGTGATGCTCCTGATAGGGTGTG 307
QY 241 C 241
DB 308 C 308

RESULT 5
AA24835

ID AAX24835 standard; DNA; 341 BP.

XX AAX24835;

XX 21-JUN-1999 (first entry)

XX Infectious hepatitis C virus genotype 1b 5'UTR.

XX HCV; infectious clone; infection; diagnosis; therapy; vaccine;
XX screening; assay; antiviral; virucide; ss.

XX Hepatitis C virus.

XX WO9904008-A2;

XX 28-JAN-1999.

XX 16-JUL-1998; 98WO-US14688.

XX 27-JAN-1998; 98US-0014416.

XX 18-JUL-1997; 97US-0053062.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Bukh J, Emerson SU, Purcell RH, Yanagi M;

XX WPI; 1999-13252/11.

XX New isolated hepatitis C virus nucleic acids - used to develop
XX products for the diagnosis, prevention and treatment of HCV
XX infections and for developing screening assays

XX Example 7; Fig 11; 126pp; English.

CC The present sequence comprises the 5' untranslated region (5'UTR)
CC of hepatitis C virus (HCV) genotype 1b infectious clone PCV-J416S.
CC This has been compared with the 5'UTR of genotype 1a infectious
CC clone PCV-H77C (see AAX24836) and a consensus (see AAX24834) produced.
CC The invention discloses nucleic acid sequences (see AAX24832-33 and
CC AAX24843) which encode infectious HCV viruses, and the use of these
CC sequences, and polypeptides (see AAW98020-22) encoded by them, in the
CC development of vaccines and diagnostics for HCV and in the
CC development of screening assays for the identification of antiviral
CC agents for HCV.

SQ Sequence 341 BP; 63 A; 103 C; 106 G; 69 T; 0 other;

Query Match 100.0%; Score 241; DB 20; Length 341;
Best Local Similarity 100.0%; Pred. No. 3e-62;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCATGGCTGTAGTATGATGTCGACCTCCAGAGACCCCTT 60
DB 68 GCAGAAAGCGCTAGCATGGCTGTAGTATGATGTCGACCTCCAGAGACCCCTT 127
QY 61 CCCGGAGAGCATTAGTGTCTGCGAACCAGGAGATACACCGGAATTGCCAGAGACC 120
DB 128 CCCGGAGAGCATTAGTGTCTGCGAACCAGGAGATACACCGGAATTGCCAGAGACC 187
QY 121 GGGTCCTTCTTGGATCAACCCGCTCAATGCTGAGATTTGGGGTGCCTCCGCGAGAC 180
DB 188 GGGTCCTTCTTGGATCAACCCGCTCAATGCTGAGATTTGGGGTGCCTCCGCGAGAC 247
QY 181 TGGTACCGAGTAGTGTGGTGGCGAAGGCGCTTGGTACTGCTGATAGAGTGTG 240
DB 248 TGGTACCGAGTAGTGTGGTGGCGAAGGCGCTTGGTACTGCTGATAGAGTGTG 307
QY 241 C 241
DB 308 C 308

RESULT 6
AA25773/C

ID AA257773 standard; DNA; 342 BP.

XX AA257773;

XX 05-APR-2000 (first entry)

XX Hepatitis C virus antisense inhibitor oligonucleotide #39.

XX Hepatitis C virus; HCV; antisense oligonucleotide; hepatotropic; ss;
XX anti-inflammatory; translation inhibition; HCV infection; virucide.

XX Hepatitis C virus.

XX US6001990-A.

XX 14-DEC-1999.

XX 07-JUN-1995; 95US-0474700.

XX 10-MAY-1994; 94US-0240382.

XX (GENO) GEN HOSPITAL CORP.

XX Moradpour D, Wands JR, Wakita T;

XX WPI; 2000-104900/09.

XX Antisense oligonucleotide to Hepatitis C virus RNA, useful for treating
XX Hepatitis C virus infections

XX Claim 28; Column 31; 31pp; English.

CC This sequence is an antisense oligonucleotide that hybridizes to
CC Hepatitis C virus (HCV) RNA, under physiological conditions. The
CC invention relates to HCV antisense oligonucleotides, and also for a
CC vector comprising a nucleotide sequence, which is transcribed in an animal
CC cell to generate an antisense oligonucleotide. The oligonucleotides have
CC virucide, hepatotropic and anti-inflammatory activity, and are useful for
CC treating HCV infection by inhibiting translation of type I-V HCV RNA.
CC Hepatitis C virus is a positive strand RNA virus, and is the major
CC causative agent of post-transfusion hepatitis. Persistent HCV infection
CC can lead to chronic hepatitis, cirrhosis, and hepatocellular carcinoma.

SQ Sequence 342 BP; 69 A; 106 C; 103 G; 64 T; 0 other;

Query Match 100.0%; Score 241; DB 21; Length 342;
Best Local Similarity 100.0%; Pred. No. 3e-62;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCATGGCTGTAGTATGATGTCGACCTCCAGAGACCCCTT 60
DB 275 GCAGAAAGCGCTAGCATGGCTGTAGTATGATGTCGACCTCCAGAGACCCCTT 216
QY 61 CCCGGAGAGCATTAGTGTCTGCGAACCAGGAGATACACCGGAATTGCCAGAGACC 120
DB 215 CCCGGAGAGCATTAGTGTCTGCGAACCAGGAGATACACCGGAATTGCCAGAGACC 156
QY 121 GGGTCCTTCTTGGATCAACCCGCTCAATGCTGAGATTTGGGGTGCCTCCGCGAGAC 180
DB 155 GGGTCCTTCTTGGATCAACCCGCTCAATGCTGAGATTTGGGGTGCCTCCGCGAGAC 96
QY 181 TGGTACCGAGTAGTGTGGTGGCGAAGGCGCTTGGTACTGCTGATAGAGTGTG 240
DB 95 TGGTACCGAGTAGTGTGGTGGCGAAGGCGCTTGGTACTGCTGATAGAGTGTG 36
QY 241 C 241
DB 35 C 35

```
RESULT 7
ABA96260
ID ABA96260 standard; DNA; 393 BP.
XX
AC ABA96260;
XX
XX
DT 15-MAR-2002 (first entry)
XX
DE Hepatitis C virus RNA polymerase encoding polynucleotide SEQ ID NO 5..
XX
KM Hepatitis C virus; RNA polymerase; RNA virus replication; ds.
XX
OS Hepatitis C virus.
XX
PN WO20018161-A1.
XX
PD 22-NOV-2001.
XX
PF 15-MAY-2001; 2001WO-JP04033.
XX
PR 15-MAY-2000; 2000JP-0142451.
XX
PA (TOKM-) TOKYO METROPOLITAN ORG MEDICAL RES.
XX
PA (CHUS ) CHUGAI SEIYAKU KK.
XX
PI Kohara M, Matsuzaki J, Okamoto K, Katsume T;
XX
PS WPI; 2002-114236/15.
XX
DR Vector for analysing RNA virus replication, contains DNA encoding RNA
XX
PT virus-derived RNA dependent RNA polymerase
XX
PS Example 1; Page 55; 66pp; Japanese.
XX
CC The invention relates to a vector for analysing RNA virus replication,
XX
CC comprising DNA encoding RNA virus-derived RNA dependent RNA polymerase
XX
CC (ABA96256-ABA96261), for evaluation of RNA virus-derived RNA dependent
XX
CC RNA polymerase activity by studying interaction between viral proteins,
XX
CC host factors and the level of expression of reporter genes.
XX
SQ Sequence 393 BP; 83 A; 115 C; 120 G; 75 T; 0 other;
XX
Query Match 100.0%; Score 241; DB 24; Length 393;
Best Local Similarity 100.0%; Pred. No. 3; le-62;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GCAGAAAGCGTCTAGCCATGCGGCTTGTAGTGTGCTGCAGCCTCCAGGACCCCCCT 60
DB 80 GCAGAAAGCGTCTAGCCATGCGGCTTGTAGTGTGCTGCAGCCTCCAGGACCCCCCT 139
OY 61 CCCGGAGAGCCATAGTGTGCGGAACCGGTAGTACACCGGAATTGCCAGAGCAGAC 120
DB 140 CCCGGAGAGCCATAGTGTGCGGAACCGGTAGTACACCGGAATTGCCAGAGCAGAC 199
OY 121 GGGTCTCTTCTTGATCAACCCGCTCAATGCTGAGATTGGGGCTCCCCCGGAGAC 180
DB 200 GGGTCTCTTCTTGATCAACCCGCTCAATGCTGAGATTGGGGCTCCCCCGGAGAC 259
OY 181 TGCTAGCCGAGTAGTGTGGTGGCGAAGGCGCTTGCTGACTGCTGATAGGGTCTTG 240
DB 260 TGCTAGCCGAGTAGTGTGGTGGCGAAGGCGCTTGCTGACTGCTGATAGGGTCTTG 319
OY 241 C 241
DB 320 C 320
DB
```

```
RESULT 8
ABA96259
ID ABA96259 standard; DNA; 412 BP.
XX
AC ABA96259;
XX
XX
DT 15-MAR-2002 (first entry)
XX
DE Hepatitis C virus RNA polymerase encoding polynucleotide SEQ ID NO 4.
XX
KM Hepatitis C virus; RNA polymerase; RNA virus replication; ds.
XX
OS Hepatitis C virus.
XX
PN WO20018161-A1.
XX
PD 22-NOV-2001.
XX
PF 15-MAY-2001; 2001WO-JP04033.
XX
PR 15-MAY-2000; 2000JP-0142451.
XX
PA (TOKM-) TOKYO METROPOLITAN ORG MEDICAL RES.
XX
PA (CHUS ) CHUGAI SEIYAKU KK.
XX
PI Kohara M, Matsuzaki J, Okamoto K, Katsume T;
XX
PS WPI; 2002-114236/15.
XX
DR Vector for analysing RNA virus replication, contains DNA encoding RNA
XX
PT virus-derived RNA dependent RNA polymerase
XX
PS Example 1; Page 55; 66pp; Japanese.
XX
CC The invention relates to a vector for analysing RNA virus replication,
XX
CC comprising DNA encoding RNA virus-derived RNA dependent RNA polymerase
XX
CC (ABA96256-ABA96261), for evaluation of RNA virus-derived RNA dependent
XX
CC RNA polymerase activity by studying interaction between viral proteins,
XX
CC host factors and the level of expression of reporter genes.
XX
SQ Sequence 412 BP; 92 A; 119 C; 120 G; 81 T; 0 other;
XX
Query Match 100.0%; Score 241; DB 24; Length 412;
Best Local Similarity 100.0%; Pred. No. 3; le-62;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GCAGAAAGCGTCTAGCCATGCGGCTTGTAGTGTGCTGCAGCCTCCAGGACCCCCCT 60
DB 99 GCAGAAAGCGTCTAGCCATGCGGCTTGTAGTGTGCTGCAGCCTCCAGGACCCCCCT 158
OY 61 CCCGGAGAGCCATAGTGTGCGGAACCGGTAGTACACCGGAATTGCCAGAGCAGAC 120
DB 159 CCCGGAGAGCCATAGTGTGCGGAACCGGTAGTACACCGGAATTGCCAGAGCAGAC 218
OY 121 GGGTCTCTTCTTGATCAACCCGCTCAATGCTGAGATTGGGGCTCCCCCGGAGAC 180
DB 219 GGGTCTCTTCTTGATCAACCCGCTCAATGCTGAGATTGGGGCTCCCCCGGAGAC 278
OY 181 TGCTAGCCGAGTAGTGTGGTGGCGAAGGCGCTTGCTGACTGCTGATAGGGTCTTG 240
DB 279 TGCTAGCCGAGTAGTGTGGTGGCGAAGGCGCTTGCTGACTGCTGATAGGGTCTTG 338
OY 241 C 241
DB 339 C 339
DB
```

```
RESULT 9
AA032444
ID AA032444 standard; DNA; 483 BP.
XX
AC AA032444;
XX
XX
DT 26-APR-1993 (first entry)
XX
DE HCV core-envelope clone NI-1.
XX
KM Clone; polypeptide; core-envelope; region; Hepatitis C virus; HCV;
XX
XX HC; cDNA; transcriptase; primer; allele; diagnostic; antibody;
```

KM suppress; control; proteolytic processing; precursor; ss.
 XX Hepatitis C virus.
 XX
 XX Key Location/Qualifiers
 FT CDS 318..482
 FT /*tag= a
 XX
 PN EP518313-A.
 XX
 PD 16-DEC-1992.
 XX
 PF 11-JUN-1992; 92EP-0109812.
 XX
 PR 11-JUN-1991; 91JP-0139268.
 PR 12-JUL-1991; 91JP-0172794.
 PR 07-OCT-1991; 91JP-0287008.
 PR 16-DEC-1991; 91JP-0332329.
 PR 20-APR-1992; 92JP-0099957.
 XX
 PA (MITU) MITSUBISHI KASEI CORP.
 XX
 PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
 PI Teranishi Y;
 XX
 DR WPI: 1992-417213/51.
 DR P-PDB: AAR29534.
 XX
 PT New hepatitis C virus gene and its encoded protein - used for
 PT diagnosing and vaccinating against hepatitis C virus infections
 XX
 PS Claim 33; Page 74; 305pp; English.
 XX
 CC The sequences given in AA032438-39 and AA032444-53 are various clones
 CC which encode the polypeptide core-envelope region of the Hepatitis C
 CC virus (HCV) gene of the invention. These sequences were isolated from
 CC the serum of a patient suffering from hepatitis C (HC). The isolated
 CC RNA sequences were converted into cDNA using transcriptase in the
 CC presence of one of the primer sequences given in AA032540-46. The cDNA
 CC sequences isolated represent different alleles of the same region of
 CC the HCV gene. The entire HCV gene (see AA032436) is useful in the
 CC development of a diagnostic method which is more accurate and
 CC effective than conventional ones. In the detection of antibodies
 CC raised against a wide range of HCVs which have been hardly detected
 CC before. The complete gene may be used in an in vitro screening system
 CC for a substance capable of specifically suppressing or controlling a
 CC proteolytic processing of a precursor polypeptide of HCV.
 CC
 XX
 SO Sequence 483 BP; 102 A; 141 C; 146 G; 94 T; 0 other;
 Query Match 100.0%; Score 241; DB 13; Length 483;
 Best Local Similarity 100.0%; Pred. No. 3.2e-62;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCGCTAGCCATGCGGTAGTATGAGTGTCTGCAAGCCCTCAGAGACCCCTT 60
 Db 44 GCAGAAAGCGCTAGCCATGCGGTAGTATGAGTGTCTGCAAGCCCTCAGAGACCCCTT 103
 Oy 61 CCGGAGAGAGCCATGAGTGTCTGCGAACCCTGAGTACACCGGAATTGCCAGAGACGACC 120
 Db 104 CCGGAGAGAGCCATGAGTGTCTGCGAACCCTGAGTACACCGGAATTGCCAGAGACGACC 163
 Oy 121 GGGTCTTTCTTGATCAACCCGCTCAATGCTGAGAGATTGGGGTGGCCCCGGGAGAC 180
 Db 164 GGGTCTTTCTTGATCAACCCGCTCAATGCTGAGAGATTGGGGTGGCCCCGGGAGAC 223
 Oy 181 TCGTAGCCGAGTAGTGTGGTGTGCGGAAGGCTTGTGATGCTGATAGAGGTGCTTG 240
 Db 224 TCGTAGCCGAGTAGTGTGGTGTGCGGAAGGCTTGTGATGCTGATAGAGGTGCTTG 283
 Oy 241 C 241
 Db 284 C 284

RESULT 10
 AA032445
 ID AA032445 standard; DNA; 483 BP.
 XX
 AC AA032445;
 XX
 DT 26-APR-1993 (first entry)
 XX
 DE HCV core-envelope clone N1-2.
 XX
 KM Clone; polypeptide; core-envelope; region; Hepatitis C; Virus; HCV;
 KM HC; cDNA; transcriptase; primer; allele; diagnostic; antibody;
 KM suppress; control; proteolytic processing; precursor; ss.
 OS Hepatitis C virus.
 XX
 XX Key Location/Qualifiers
 FT CDS 318..482
 FT /*tag= a
 XX
 PN EP518313-A.
 XX
 PD 16-DEC-1992.
 XX
 PF 11-JUN-1992; 92EP-0109812.
 XX
 PR 11-JUN-1991; 91JP-0139268.
 PR 12-JUL-1991; 91JP-0172794.
 PR 07-OCT-1991; 91JP-0287008.
 PR 16-DEC-1991; 91JP-0332329.
 PR 20-APR-1992; 92JP-0099957.
 XX
 PA (MITU) MITSUBISHI KASEI CORP.
 XX
 PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
 PI Teranishi Y;
 XX
 DR WPI: 1992-417213/51.
 DR P-PDB: AAR29535.
 XX
 PT New hepatitis C virus gene and its encoded protein - used for
 PT diagnosing and vaccinating against hepatitis C virus infections
 XX
 PS Claim 33; Page 87; 305pp; English.
 XX
 CC The sequences given in AA032438-39 and AA032444-53 are various clones
 CC which encode the polypeptide core-envelope region of the Hepatitis C
 CC virus (HCV) gene of the invention. These sequences were isolated from
 CC the serum of a patient suffering from hepatitis C (HC). The isolated
 CC RNA sequences were converted into cDNA using transcriptase in the
 CC presence of one of the primer sequences given in AA032540-46. The cDNA
 CC sequences isolated represent different alleles of the same region of
 CC the HCV gene. The entire HCV gene (see AA032436) is useful in the
 CC development of a diagnostic method which is more accurate and
 CC effective than conventional ones. In the detection of antibodies
 CC raised against a wide range of HCVs which have been hardly detected
 CC before. The complete gene may be used in an in vitro screening system
 CC for a substance capable of specifically suppressing or controlling a
 CC proteolytic processing of a precursor polypeptide of HCV.
 CC
 XX
 SO Sequence 483 BP; 102 A; 141 C; 146 G; 94 T; 0 other;
 Query Match 100.0%; Score 241; DB 13; Length 483;
 Best Local Similarity 100.0%; Pred. No. 3.2e-62;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCGCTAGCCATGCGGTAGTATGAGTGTCTGCAAGCCCTCAGAGACCCCTT 60
 Db 44 GCAGAAAGCGCTAGCCATGCGGTAGTATGAGTGTCTGCAAGCCCTCAGAGACCCCTT 103
 Oy 61 CCGGAGAGAGCCATGAGTGTCTGCGAACCCTGAGTACACCGGAATTGCCAGAGACGACC 120

Best Local Similarity 79.3%; Pred. No. 3.4e-62;
Matches 191; Conservative 50; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTTCAGCCATGGCGTGTAGTGTGCTGTCGACGCTCCAGACCCCT 60
DB 68 GCAGAAAGCGTTCAGCCATGGCGTGTAGTGTGCTGTCGACGCTCCAGACCCCT 127
OY 61 CCCGGAGAGCCATAGTGTGTCGCGAACCGGTGAGTACACCGGATTCGCGAGAC 120
DB 128 CCCGGAGAGCCATAGTGTGTCGCGAACCGGTGAGTACACCGGATTCGCGAGAC 187
OY 121 GGGTCTTTCTTGGATGATCAACCCGCTCATGCTGAGATTGGGCGTCCCGGAGAC 180
DB 188 GGGTCTTTCTTGGATGATGATCAACCCGCTCATGCTGAGATTGGGCGTCCCGGAGAC 247
OY 181 TGCCTAGCCGAGTGTGTTGGTGGCGGAAAGCGCTTGTGCTGATAGGCTCTG 240
DB 248 TGCCTAGCCGAGTGTGTTGGTGGCGGAAAGCGCTTGTGCTGATAGGCTCTG 307
OY 241 C 241
DB 308 C 308

RESULT 13.

AAZ57396
ID AAZ57396 standard; DNA; 703 BP.

XX AAZ57396;

DT 07-APR-2000 (first entry)

DE Hepatitis C virus DNA sequence SMO ID NO:3.

KW Hepatitis C virus; RNA virus; replication; viral infection; ds.

OS Hepatitis C virus.

PN MO9967394-A1.

PD 29-DEC-1999.

PF 24-JUN-1999; 99MO-JP03380.

PR 24-JUN-1998; 98JP-0177820.

PA (CHUS) CHUGAI SEIYAKU KK.

PI Kohara M, Kohara K, Taira K, Matsuzaki J, Ohmori H;

DR WPI; 2000-106296/09.

PT Vectors expressing full-length gene of RNA viruses, useful in
clarifying mechanisms of RNA viral replication, infection, and
developing remedies and therapeutics

PS Example 1; Page 35-36; 46pp; Japanese.

CC The present invention describes a vector comprising a cDNA encoding an
RNA virus gene, constructed to ensure the exact and homogeneous
transcription of both terminals of the RNA virus gene. Also described
is a method for screening drugs for inhibiting the replication of RNA
virus by using the RNA viral infection model animal, particularly one
with hepatitis C viral infection. The vector is useful in clarifying
the mechanism of RNA viral replication, onset of RNA viral infection,
and developing remedies and therapeutics for RNA viral infection,
particularly of a hepatitis C virus. The present sequence represents
a DNA sequence from hepatitis C virus from an example of the present
invention.

Sequence 703 BP; 151 A; 211 C; 213 G; 128 T; 0 other;

Query Match 100.0%; Score 241; DB 21; Length 703;

Best Local Similarity 100.0%; Pred. No. 3.4e-62;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTTCAGCCATGGCGTGTAGTGTGCTGTCGACGCTCCAGACCCCT 60
DB 158 GCAGAAAGCGTTCAGCCATGGCGTGTAGTGTGCTGTCGACGCTCCAGACCCCT 217
OY 61 CCCGGAGAGCCATAGTGTGTCGCGAACCGGTGAGTACACCGGATTCGCGAGAC 120
DB 218 CCCGGAGAGCCATAGTGTGTCGCGAACCGGTGAGTACACCGGATTCGCGAGAC 277
OY 121 GGGTCTTTCTTGGATGATCAACCCGCTCATGCTGAGATTGGGCGTCCCGGAGAC 180
DB 278 GGGTCTTTCTTGGATGATCAACCCGCTCATGCTGAGATTGGGCGTCCCGGAGAC 337
OY 181 TGCCTAGCCGAGTGTGTTGGTGGCGGAAAGCGCTTGTGCTGATAGGCTCTG 240
DB 338 TGCCTAGCCGAGTGTGTTGGTGGCGGAAAGCGCTTGTGCTGATAGGCTCTG 397
OY 241 C 241
DB 398 C 398

RESULT 14

AAA08097
ID AAA08097 standard; cDNA; 713 BP.

XX AAA08097;

DT 22-JUN-2000 (first entry)

DE Hepatitis type C virus nucleotide sequence SEQ ID NO:1.

KW Hepatitis C virus; HCV; 5' UTR; 5'-nontranslational region; diagnosis;

KM gene expression; infection; IRES; viral; ss.

OS Hepatitis C virus.

FN Key Location/Qualifiers

FT 5'UTR 1..341

FT CDS /*tag= a

FT 342..713

FT /*tag= b

FT /note= "no stop codon given"

PN MO200012691-A1.

PD 09-MAR-2000.

PF 08-JUL-1999; 99MO-JP03682.

PR 27-AUG-1998; 98JP-0241367.

PA (FUSO) FUSO PHARM IND LTD.

PI Yamada O, Yoshida H, Zhang J;

DR WPI; 2000-237867/20.

PS P-PSDB; AAY82358.

CC Claim 18; Page 82-84; 94pp; Japanese.

CC The present invention describes an oligonucleotide sequence which
potentiates the expression of a useful gene when incorporated in a
gene expression vector, where the oligonucleotide sequence consists
of the 5'-non-translated region (5'UTR) of a viral genome or its
fragment or modified form. The 5'UTR sequence is useful in screening
potential initiation factors of interaction with the viral IRES or

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 285.951 seconds

(without alignments)
3358.597 Million cell updates/sec

Title: US-10-087-631B-3

Perfect score: 33

Sequence: 1 cgggtactaccgctccgcagacacatgac 33

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: GenBank:*
2: gb_ba:*
3: gb_htg:*
4: gb_in:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_higo_hum:*
40: em_higo_mus:*
41: em_higo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	30.4	92.1	36	6	A52660	A52660 Sequence 5
C 2	27.4	83.0	29	6	A68291	A68291 Sequence 12
C 3	22	66.7	68	6	AX003355	AX003355 Sequence
C 4	22	66.7	68	6	AX003356	AX003356 Sequence
C 5	22	66.7	116	14	HPCHAI12	M74254 Hepatitis C
C 6	22	66.7	116	14	HPCHAI13	M74255 Hepatitis C
C 7	22	66.7	125	14	HPCHAI10	M74252 Hepatitis C
C 8	22	66.7	140	6	E10301	E10301 Anti-sense
C 9	22	66.7	142	14	S72378	S72378 (5' region)
C 10	22	66.7	148	14	HPCHAI6	M74248 Hepatitis C
C 11	22	66.7	148	14	S75838	S75838 NS-5 (5' un
C 12	22	66.7	149	14	HPCEB1	D10114 Hepatitis C
C 13	22	66.7	149	14	HPCEB11	D10119 Hepatitis C
C 14	22	66.7	149	14	HPCEB12	D10119 Hepatitis C
C 15	22	66.7	149	14	HPCEB13	D10120 Hepatitis C
C 16	22	66.7	149	14	HPCEB16	D10121 Hepatitis C
C 17	22	66.7	149	14	HPCEB18	D10122 Hepatitis C
C 18	22	66.7	149	14	HPCEB2	D10113 Hepatitis C
C 19	22	66.7	149	14	HPCEB7	D10115 Hepatitis C
C 20	22	66.7	149	14	HPCEB8	D10116 Hepatitis C
C 21	22	66.7	149	14	HPCEB9	D10117 Hepatitis C
C 22	22	66.7	155	14	HC056523	U56523 Hepatitis C
C 23	22	66.7	155	14	HC056524	U56524 Hepatitis C
C 24	22	66.7	155	14	HC056525	U56525 Hepatitis C
C 25	22	66.7	155	14	HC056526	U56526 Hepatitis C
C 26	22	66.7	155	14	HC056527	U56527 Hepatitis C
C 27	22	66.7	155	14	HC056528	U56528 Hepatitis C
C 28	22	66.7	155	14	HC056529	U56529 Hepatitis C
C 29	22	66.7	155	14	HC056530	U56530 Hepatitis C
C 30	22	66.7	155	14	HC056531	U56531 Hepatitis C
C 31	22	66.7	155	14	HC056532	U56532 Hepatitis C
C 32	22	66.7	155	14	HC056533	U56533 Hepatitis C
C 33	22	66.7	155	14	HC056534	U56534 Hepatitis C
C 34	22	66.7	155	14	HC056535	U56535 Hepatitis C
C 35	22	66.7	155	14	HC056536	U56536 Hepatitis C
C 36	22	66.7	155	14	HC056537	U56537 Hepatitis C
C 37	22	66.7	155	14	HC056538	U56538 Hepatitis C
C 38	22	66.7	155	14	HC056539	U56539 Hepatitis C
C 39	22	66.7	155	14	HC056540	U56540 Hepatitis C
C 40	22	66.7	155	14	HC056541	U56541 Hepatitis C
C 41	22	66.7	155	14	HC056542	U56542 Hepatitis C
C 42	22	66.7	155	14	HC056543	U56543 Hepatitis C
C 43	22	66.7	155	14	HC056544	U56544 Hepatitis C
C 44	22	66.7	155	14	HC056545	U56545 Hepatitis C
C 45	22	66.7	155	14	HC056546	U56546 Hepatitis C

ALIGNMENTS

RESULT 1
LOCUS A52660 36 bp
DEFINITION Sequence 5 from Patent WO9624662.
ACCESSION A52660
VERSION A52660.1 GI:2851824
KEYWORDS
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 36)
AUTHORS Ravagnan,G., Battaglia,M., Carloni,G., Ponzelto,A. and Iacovacci,S.
TITLE PROCESS TO 'IN VITRO' PROPAGATE THE HEPATITIS C VIRUS (HCV) IN NON
JOURNAL LYMPHOBLASTOID ANIMAL CELL CULTURES AND PRODUCTS THEREOF
Patent: WO 9624662-A 5 15-AUG-1996;

Pred. No. is the number of results predicted by chance to have a

COMMENT
CONSIGLIO NAZIONALE RICERCA (IT)
Other publication AU 1822195 960827.
FEATURES
Location/Qualifiers
1. 36
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT
9 a 9 c 11 g 7 t
ORIGIN

Query Match
Best Local Similarity 92.1%; Score 30.4; DB 6; Length 36;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y
1 CGGTGTTACTCACCCTTCGCGAGACCACTATGG 32
32 CGGTGTTACTCACCCTTCGCGAGACCACTATGG 1

RESULT 2
A68291/c
LOCUS A68291. Sequence 12 from Patent WO9746716. 29 bp DNA linear PAT 06-MAY-1999
DEFINITION
ACCESSION A68291
VERSION A68291.1 GI:4759412
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE
1 (bases 1 to 29)
Bosio, P., Strumia, C. and Clementza, F.
METHOD TO DETECT HCV SPECIFIC NUCLEIC ACIDS
TITLE
JOURNAL Patent: NO 9746716-A 12 11-DEC-1997;
WABCO B V (NL)
Other publication IT RM960404 19971209.
COMMENT
FEATURES
Location/Qualifiers
1. 29
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT
7 a 7 c 9 g 6 t
ORIGIN

Query Match
Best Local Similarity 83.0%; Score 27.4; DB 6; Length 29;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y
4 TGTACTCACCCTTCGCGAGACCACTATGG 32
29 TGTACTCACCCTTCGCGAGACCACTATGG 1

RESULT 3
AX003355/c
LOCUS AX003355. 68 bp DNA linear PAT 24-AUG-2000
DEFINITION
ACCESSION AX003355
VERSION AX003355.1 GI:9927160
KEYWORDS
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
REFERENCE
1 (bases 1 to 68)
Priml, D. and Mantero, G.
METHODS OF DETECTING POLYNUCLEOTIDE ANALYTES
TITLE
JOURNAL Patent: WO 9928503-A 6 10-JUN-1999;
PRIMI DANIELE (IT); MANTERO GIOVANNI (IT)
Location/Qualifiers
1. 68
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT
14 a 17 c 24 g 13 t
ORIGIN

Query Match
Best Local Similarity 66.7%; Score 22; DB 6; Length 68;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Y
1 CGGTGTTACTCACCCTTCGCGAGACCACTATGGC 33
40 CGGTGTTACTCACCCTTCGCGAGACCACTATGGC 7

RESULT 4
AX003356
LOCUS AX003356. 68 bp DNA linear PAT 24-AUG-2000
DEFINITION
ACCESSION AX003356
VERSION AX003356.1 GI:9927161
KEYWORDS
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
REFERENCE
1 (bases 1 to 68)
Priml, D. and Mantero, G.
METHODS OF DETECTING POLYNUCLEOTIDE ANALYTES
TITLE
JOURNAL Patent: WO 9928503-A 7 10-JUN-1999;
PRIMI DANIELE (IT); MANTERO GIOVANNI (IT)
Location/Qualifiers
1. 68
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT
13 a 24 c 17 g 14 t
ORIGIN

Query Match
Best Local Similarity 66.7%; Score 22; DB 6; Length 68;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Y
1 CGGTGTTACTCACCCTTCGCGAGACCACTATGGC 33
29 CGGTGTTACTCACCCTTCGCGAGACCACTATGGC 62

RESULT 5
HPCCHA12/c
LOCUS HPCCHA12. 116 bp ss-RNA linear VRL 02-AUG-1993
DEFINITION
ACCESSION M74254
VERSION M74254.1 GI:329743
KEYWORDS
SOURCE Hepatitis C virus cDNA to genomic RNA.
ORGANISM Hepatitis C virus
REFERENCE
1 (bases 1 to 116)
Cha, T.-A., Kolberg, J., Irvine, B., Stempien, M., Beall, E., Yano, M.,
Choo, Q.-L., Houghton, M., Ruo, G., Han, J. H. and Ueda, M. S.
Use of a signature nucleotide sequence of the hepatitis C virus for
the detection of viral RNA in human serum and plasma
Unpublished (1991)
Location/Qualifiers
1. 116
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT
23 a 40 c 33 g 20 t
ORIGIN
Map position 108.

Query Match
Best Local Similarity 66.7%; Score 22; DB 14; Length 116;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Y
1 CGGTGTTACTCACCCTTCGCGAGACCACTATGGC 33
63 CGGTGTTACTCACCCTTCGCGAGACCACTATGGC 30

```

RESULT 6
LOCUS HPCCHA13 116 bp ss-RNA linear VRL 02-AUG-1993
DEFINITION Hepatitis C virus isolate 1128 genome, 5' untranslated region.
ACCESSION M74255
VERSION M74255.1 GI:329744
KEYWORDS
SOURCE
ORGANISM Hepatitis C virus cDNA to genomic RNA.
Virus; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE 1 (bases 1 to 116)
AUTHORS Cha,T.-A., Kolberg,J., Irvine,B., Stempien,M., Beall,E., Yano,M.,
Choo,Q.-L., Houghton,M., Kuo,G., Han,J.H. and Urdea,M.S.
TITLE Use of a signature nucleotide sequence of the hepatitis C virus for
the detection of viral RNA in human serum and plasma
JOURNAL Unpublished (1991)
FEATURES
source
Location/Qualifiers
1..116
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT 23 a 40 c 33 g 20 t
ORIGIN
Map position 108.

Query Match 66.7%; Score 22; DB 14; Length 116;
Best Local Similarity 97.1%; Pred. No. 4.8e+02;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTACAC-GTTCGCGAGACCACTATGCG 33
Db 63 CGGTGTACTACACGCGTTCGCGAGACCACTATGCG 30

RESULT 7
LOCUS HPCCHA10 125 bp ss-RNA linear VRL 02-AUG-1993
DEFINITION Hepatitis C virus isolate 1124 genome, 5' untranslated region.
ACCESSION M74252
VERSION M74252.1 GI:329741
KEYWORDS
SOURCE
ORGANISM Hepatitis C virus cDNA to genomic RNA.
Virus; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE 1 (bases 1 to 125)
AUTHORS Cha,T.-A., Kolberg,J., Irvine,B., Stempien,M., Beall,E., Yano,M.,
Choo,Q.-L., Houghton,M., Kuo,G., Han,J.H. and Urdea,M.S.
TITLE Use of a signature nucleotide sequence of the hepatitis C virus for
the detection of viral RNA in human serum and plasma
JOURNAL Unpublished (1991)
FEATURES
source
Location/Qualifiers
1..125
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT 24 a 42 c 37 g 22 t
ORIGIN
Map position 99.

Query Match 66.7%; Score 22; DB 14; Length 125;
Best Local Similarity 97.1%; Pred. No. 4.8e+02;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTACAC-GTTCGCGAGACCACTATGCG 33
Db 72 CGGTGTACTACACGCGTTCGCGAGACCACTATGCG 39

RESULT 8
LOCUS E10301 140 bp RNA linear PAT 29-SEP-1997
DEFINITION Anti-sense RNA against partial sequence of 5'UTR of hepatitis C
virus.
ACCESSION E10301

```

```

VERSION E10301.1 GI:5708588
KEYWORDS JP 1995303485-A/2.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 140)
AUTHORS Funahashi,S. and Hasegawa,A.
TITLE HCV ANTI-SENSE RNA, MANIFESTATION VECTOR CONTAINING THE SAME, AND
THERAPEUTIC AGENT FOR HCV-INVOLVED DISEASE CONTAINING THE RNA OR
MANIFESTATION VECTOR
JOURNAL Patent: JP 1995303485-A 2 21-NOV-1995;
TONEIN CORP
OS None
OC Artificial sequences.
PN JP 1995303485-A/2
PD 21-NOV-1995
PE 13-MAY-1994 JP 1994124609
PI FUNAHASHI SHINICHI, HASEGAWA AKIRA
PC C12N15/09,A61K31/70,A61K31/70,C07H21/04;
CC strandedness: Single;
CC topology: linear;
CC hypothetical: No;
CC anti-sense: Yes;
CC key Location/Qualifiers
FH source 1..140
FT /organism="Artificial sequences" FT
FT misc-feature 1..140 /note="Anti-sense RNA against partial sequence
of 5'UTR of
hepatitis C virus".
FT Location/Qualifiers
1..140
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 26 a 43 c 46 g 25 t
ORIGIN

Query Match 66.7%; Score 22; DB 6; Length 140;
Best Local Similarity 97.1%; Pred. No. 4.7e+02;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTACAC-GTTCGCGAGACCACTATGCG 33
Db 80 CGGTGTACTACACGCGTTCGCGAGACCACTATGCG 113

RESULT 9
LOCUS S72378 142 bp RNA linear VRL 24-JAN-1995
DEFINITION [5' region] [hepatitis C virus HCV, host-human liver, Genomic RNA,
142 nt].
ACCESSION S72378
VERSION S72378.1 GI:632885
KEYWORDS
SOURCE
ORGANISM Hepatitis C virus host-human liver.
Virus; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE 1 (bases 1 to 142)
AUTHORS Sullivan,D.E. and Gerber,M.A.
TITLE Conservation of hepatitis C virus 5' untranslated sequences in
hepatocellular carcinoma and the surrounding liver.
JOURNAL Hepatology 19 (3), 551-553 (1994)
MEDLINE 94164633
PUBMED 8119678
REMARK Genbank staff at the National Library of Medicine created this
entry [NCBI g1bseq 152236] from the original journal article.
This sequence comes from Fig. 1.
FEATURES
source
Location/Qualifiers
1..142
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

```

BASE COUNT

30 a 41 c 43 g 28 t

gene

/db_xref="taxon:41856"
1.148
/partial
/gene="NS-5"

Query Match
Best Local Similarity 97.1%; Score 22; DB 14; Length 142;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

BASE COUNT 29 a 40 c 50 g 29 t

OY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 33
|||||
106 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 73
|||||

Query Match
Best Local Similarity 97.1%; Score 22; DB 14; Length 148;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 33
|||||
39 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 6
|||||

RESULT 10
LOCUS HPCCH6 148 bp ss-RNA linear VRL 02-NDG-1993
DEFINITION Hepatitis C virus isolate NB011 genome, 5' untranslated region.
ACCESSION M74248
VERSION M74248.1 GI:329749
KEYWORDS
SOURCE Hepatitis C virus cDNA to genomic RNA.
ORGANISM Hepatitis C virus
VIRUSES; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

RESULT 12
LOCUS HPCEB1 149 bp RNA linear VRL 02-FEB-1999
DEFINITION Hepatitis C virus (HCV) genomic RNA, 5' non-coding region.
ACCESSION D10114
VERSION D10114.1 GI:221549
KEYWORDS

REFERENCE
AUTHORS Cha, F.-A., Kolberg, T., Irvine, B., Stempien, M., Beall, E., Yano, M., Choo, Q.-L., Houghton, M., Kuo, G., Han, J. H. and Urdea, M. S.
TITLE Use of a signature nucleotide sequence of the hepatitis C virus for the detection of viral RNA in human serum and plasma
JOURNAL Unpublished (1991)
FEATURES
SOURCE Location/Qualifiers
1.148
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

REFERENCE
AUTHORS Chan, S.W., McOmish, F., Holmes, E.C., Dow, B., Peutherer, J.F., Follett, E., Yap, P.L. and Simmonds, P.
TITLE Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants
J. Gen. Virol. 73 (Pt 5), 1131-1141 (1992)
92268871

BASE COUNT 27 a 47 c 44 g 30 t
ORIGIN Map position 76.

Query Match
Best Local Similarity 97.1%; Score 22; DB 14; Length 148;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 33
|||||
95 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 62
|||||

COMMENT

RESULT 11
LOCUS S75838 148 bp RNA linear VRL 27-JUL-1995
DEFINITION NS-5 (5' untranslated region) [hepatitis C virus 1 HCV 1, isolate A1 8, host-human, genomic RNA, 148 nt].
ACCESSION S75838
VERSION S75838.1 GI:913831
KEYWORDS
SOURCE Hepatitis C virus type 1 host-human.
ORGANISM Hepatitis C virus type 1
VIRUSES; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

FEATURES
SOURCE Location/Qualifiers
1.149
/organism="Hepatitis C virus"
/isolate="E-b1"
/db_xref="taxon:11103"

REFERENCE
AUTHORS Gerotto, M., Pontisso, P., Giostira, F., Francesconi, R., Muratori, L., Ballardini, G., Lenzi, M., Tisminețky, S., Bianchi, F. B., Baralle, F. B., et al.

TITLE Analysis of the hepatitis C virus genome in patients with anti-ILM-1 autoantibodies
J. Hepatol. 21 (2), 273-276 (1994)
JOURNAL MEDLINE
PUBMED 95081562
7989722

REMARK Genbank staff at the National Library of Medicine created this entry [NCBI gisbseq 161727] from the original journal article.
This sequence comes from Fig. 2.

FEATURES
SOURCE Location/Qualifiers
1.148
/organism="Hepatitis C virus type 1"
/isolate="A1 8"

BASE COUNT 27 a 48 c 47 g 27 t
ORIGIN

Query Match
Best Local Similarity 97.1%; Score 22; DB 14; Length 149;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 33
|||||
74 CGGTGTACTCACC-GTTCGCGAGACCACTATGCG 41
|||||

RESULT 13
LOCUS HPCEB11 149 bp RNA linear VRL 02-FEB-1999
DEFINITION Hepatitis C virus (HCV) genomic RNA, 5' non-coding region.

ACCESSION D10118
VERSION GI:221550
KEYWORDS Hepatitis C virus (isolate:E-b12) CDNA to genomic RNA.
SOURCE Hepatitis C virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
REFERENCE 1 (bases 1 to 149)
AUTHORS Chan,S.W., Mcomish,F., Holmes,E.C., Dow,B., Peutherer,J.F., Follett,E., Yap,P.L. and Simmonds,P.
TITLE Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants
JOURNAL J. Gen. Virol. 73 (Pt 5), 1131-1141 (1992)
MEDLINE 92268871
REFERENCE 2 (bases 1 to 149)
AUTHORS Chan,S.-W.
TITLE Direct Submission
JOURNAL Submitted (25-DEC-1991) Shiu-Wan Chan, University of Edinburgh, Medical School, Dept. of Medical Microbiology, Teviot Place, Edinburgh EH8 9AG U.K. (E-mail: ECHCASTLE.ED.AC.UK, Tel:031-650-3138, Fax:031-662-4135)
COMMENT Submitted (25-Dec-1991) to DDBJ by: Shiu-Wan Chan
Department of Medical Microbiology
University of Edinburgh
Medical School
Teviot Place
Edinburgh EH8 9AG
U.K.
E mail: ECHCASTLE.ED.AC.UK
Phone: 031-650-3138
Fax: 031-662-4135.
Location/Qualifiers
1. 149
/organism="Hepatitis C virus"
/isolate="E-b12"
/db_xref="taxon:11103"
31 t
BASE COUNT 29 a 47 c 42 g 31 t
ORIGIN

Query Match 66.7%; Score 22; DB 14; Length 149;
Best Local Similarity 97.1%; Pred. No. 4.7e+02;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 CGGTGACTCACC-GTTCGCGACGACCTATGGC 33
|||||
74 CGGTGACTCACC-GTTCGCGACGACCTATGGC 41
Db

RESULT 14
LOCUS HPCB12/c 149 bp RNA linear VRL 02-FEB-1999
DEFINITION Hepatitis C virus (HCV) genomic RNA, 5' non-coding region.
ACCESSION D10119
VERSION D10119.1 GI:221551
KEYWORDS Hepatitis C virus (isolate:E-b12) CDNA to genomic RNA.
SOURCE Hepatitis C virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
REFERENCE 1 (bases 1 to 149)
AUTHORS Chan,S.W., Mcomish,F., Holmes,E.C., Dow,B., Peutherer,J.F., Follett,E., Yap,P.L. and Simmonds,P.
TITLE Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants
JOURNAL J. Gen. Virol. 73 (Pt 5), 1131-1141 (1992)
MEDLINE 92268871
REFERENCE 2 (bases 1 to 149)
AUTHORS Chan,S.-W.
TITLE Direct Submission
JOURNAL Submitted (25-DEC-1991) Shiu-Wan Chan, University of Edinburgh, Medical School, Dept. of Medical Microbiology, Teviot Place, Edinburgh EH8 9AG U.K. (E-mail: ECHCASTLE.ED.AC.UK,

COMMENT Tel:031-650-3138, Fax:031-662-4135)
Submitted (25-Dec-1991) to DDBJ by:
Shiu-Wan Chan
Department of Medical Microbiology
University of Edinburgh
Medical School
Teviot Place
Edinburgh EH8 9AG
U.K.
E mail: ECHCASTLE.ED.AC.UK
Phone: 031-650-3138
Fax: 031-662-4135.
Location/Qualifiers
1. 149
/organism="Hepatitis C virus"
/isolate="E-b12"
/db_xref="taxon:11103"
31 t
BASE COUNT 28 a 47 c 43 g 31 t
ORIGIN

Query Match 66.7%; Score 22; DB 14; Length 149;
Best Local Similarity 97.1%; Pred. No. 4.7e+02;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 CGGTGACTCACC-GTTCGCGACGACCTATGGC 33
|||||
74 CGGTGACTCACC-GTTCGCGACGACCTATGGC 41
Db

RESULT 15
LOCUS HPCB13/c 149 bp RNA linear VRL 02-FEB-1999
DEFINITION Hepatitis C virus (HCV) genomic RNA, 5' non-coding region.
ACCESSION D10120
VERSION D10120.1 GI:221554
KEYWORDS Hepatitis C virus (isolate:E-b13) CDNA to genomic RNA.
SOURCE Hepatitis C virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
REFERENCE 1 (bases 1 to 149)
AUTHORS Chan,S.W., Mcomish,F., Holmes,E.C., Dow,B., Peutherer,J.F., Follett,E., Yap,P.L. and Simmonds,P.
TITLE Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants
JOURNAL J. Gen. Virol. 73 (Pt 5), 1131-1141 (1992)
MEDLINE 92268871
REFERENCE 2 (bases 1 to 149)
AUTHORS Chan,S.-W.
TITLE Direct Submission
JOURNAL Submitted (25-DEC-1991) Shiu-Wan Chan, University of Edinburgh, Medical School, Dept. of Medical Microbiology, Teviot Place, Edinburgh EH8 9AG U.K. (E-mail: ECHCASTLE.ED.AC.UK, Tel:031-650-3138, Fax:031-662-4135)
COMMENT Submitted (25-Dec-1991) to DDBJ by: Shiu-Wan Chan
Department of Medical Microbiology
University of Edinburgh
Medical School
Teviot Place
Edinburgh EH8 9AG
U.K.
E mail: ECHCASTLE.ED.AC.UK
Phone: 031-650-3138
Fax: 031-662-4135.
Location/Qualifiers
1. 149
/organism="Hepatitis C virus"
/isolate="E-b13"
/db_xref="taxon:11103"
28 t
BASE COUNT 27 a 48 c 46 g 28 t
ORIGIN

Fri Jul 11 15:29:33 2003

us-10-087-631b-3.rge

Query Match 66.7% Score 22; DB 14; Length 149;
Best Local Similarity 97.1% Pred. No. 4.7e+02;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGACTACACC-GTTCGCGAGACCACTATGGC 33
Db 74 CGGTGACTACACCGGTTCGCGAGACCACTATGGC 41

Search completed: July 10, 2003, 19:20:01
Job time : 286.951 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 242.625 Seconds

(without alignments)
3358.597 Million cell updates/sec

Title: US-10-087-631B-7

Perfect score: 28

Sequence: 1 gcaagcaccatcatcagcagcaccacaa 28

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_da:*

2: gb_hcg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_da:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hcg_hum:*

31: em_hcg_inv:*

32: em_hcg_mus:*

33: em_hcg_other:*

34: em_hcg_pln:*

35: em_hcg_rtd:*

36: em_hcg_mam:*

37: em_hcg_vrt:*

38: em_sy:*

39: em_hgtgo_hum:*

40: em_hgtgo_mus:*

41: em_hgtgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	28	AR054576	AR054576 Sequence
2	28	100.0	28	AR094138	AR094138 Sequence
3	28	100.0	28	AX147022	AX147022 Sequence
4	28	100.0	47	AR204716	AR204716 Sequence
5	28	100.0	47	AX284180	AX284180 Sequence
6	28	100.0	57	I73305	I73305 Sequence 36
7	28	100.0	75	AX021668	AX021668 Sequence
8	28	100.0	186	E10302	E10302 Anti-sense
9	28	100.0	187	AX172757	AX172757 Sequence
10	28	100.0	191	AF041266	AF041266 Hepatitis
11	28	100.0	191	AF041273	AF041273 Hepatitis
12	28	100.0	191	AF041276	AF041276 Hepatitis
13	28	100.0	191	AF041277	AF041277 Hepatitis
14	28	100.0	191	AF041279	AF041279 Hepatitis
15	28	100.0	191	AF041289	AF041289 Hepatitis
16	28	100.0	191	AF041292	AF041292 Hepatitis
17	28	100.0	191	AF041293	AF041293 Hepatitis
18	28	100.0	191	AF041294	AF041294 Hepatitis
19	28	100.0	191	AF041296	AF041296 Hepatitis
20	28	100.0	191	AF041298	AF041298 Hepatitis
21	28	100.0	191	AF041299	AF041299 Hepatitis
22	28	100.0	191	AF041300	AF041300 Hepatitis
23	28	100.0	191	AF041301	AF041301 Hepatitis
24	28	100.0	191	AF041304	AF041304 Hepatitis
25	28	100.0	191	AF041306	AF041306 Hepatitis
26	28	100.0	191	AF041308	AF041308 Hepatitis
27	28	100.0	191	AF041309	AF041309 Hepatitis
28	28	100.0	191	AF041310	AF041310 Hepatitis
29	28	100.0	191	AF041312	AF041312 Hepatitis
30	28	100.0	191	AF041313	AF041313 Hepatitis
31	28	100.0	197	HCV6315	AJ006311 Hepatitis
32	28	100.0	204	HCV6311	AJ006311 Hepatitis
33	28	100.0	204	HCV6312	AJ006312 Hepatitis
34	28	100.0	205	HCV6313	AJ006313 Hepatitis
35	28	100.0	219	AF506662	AF506662 Hepatitis
36	28	100.0	222	AF387732	AF387732 Hepatitis
37	28	100.0	222	AF387733	AF387733 Hepatitis
38	28	100.0	222	AF506641	AF506641 Hepatitis
39	28	100.0	222	AF506663	AF506663 Hepatitis
40	28	100.0	223	AF506624	AF506624 Hepatitis
41	28	100.0	224	AF506630	AF506630 Hepatitis
42	28	100.0	224	AF506635	AF506635 Hepatitis
43	28	100.0	226	AF506674	AF506674 Hepatitis
44	28	100.0	227	HCV6327	AJ006327 Hepatitis
45	28	100.0	228	AF506654	AF506654 Hepatitis

ALIGNMENTS

RESULT 1

LOCUS AR054576 28 bp DNA

DEFINITION Sequence 2 from patent US 5837442.

ACCESSION AR054576

VERSION AR054576.1 GI:580153

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 28)

AUTHORS Tsang,S.Yeo.

TITLE Oligonucleotide primers for amplifying HCV nucleic acid

JOURNAL Patent: US 5837442-A 2 17-NOV-1998;

FEATURES Location/Qualifiers

SOURCE 1.28
BASE COUNT 10 a 10 c 5 g 3 t
ORIGIN

Query Match 100.0%; Score 28; DB 6; Length 28;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGACCCCTATCAGGAGTACCA 28
DB 1 GCAGACCCCTATCAGGAGTACCA 28

RESULT 2
AR094138
LOCUS AR094138 28 bp DNA linear PAT 08-SEP-2000
DEFINITION Sequence 4 from patent US 6001611.
ACCESSION AR094138
VERSION AR094138.1 GI:10020883
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 28)
TITLE Will, S.Gordon.
JOURNAL Modified nucleic acid amplification primers.
FEATURES Patent: US 6001611-A 14-DEC-1999;
Location/Qualifiers
1.28
SOURCE /organism="unknown"

BASE COUNT 10 a 10 c 5 g 3 t
ORIGIN

Query Match 100.0%; Score 28; DB 6; Length 28;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGACCCCTATCAGGAGTACCA 28
DB 1 GCAGACCCCTATCAGGAGTACCA 28

RESULT 3
AX147022
LOCUS AX147022 28 bp DNA linear PAT 08-JUN-2001
DEFINITION Sequence 16 from Patent W00137291.
ACCESSION AX147022
VERSION AX147022.1 GI:14346293
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 28)
AUTHORS Weindel, K., Riedling, M. and Geiger, A.
TITLE Magnetic glass particles, method for their preparation and uses thereof
JOURNAL Patent: WO 0137291-A 16-25-MAY-2001;
FEATURES Roche Diagnostics GmbH (DE)
Location/Qualifiers
1.28
SOURCE /organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic oligonucleotide (HCV reverse)"

modified_base 28
BASE COUNT 10 a 10 c 5 g 3 t
ORIGIN

Query Match 100.0%; Score 28; DB 6; Length 28;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGACCCCTATCAGGAGTACCA 28
DB 1 GCAGACCCCTATCAGGAGTACCA 28

RESULT 4
AR204716/c
LOCUS AR204716 47 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6368801.
ACCESSION AR204716
VERSION AR204716.1 GI:21502113
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS Faruqi, A.Fawad.
TITLE Detection and amplification of RNA using target-mediated ligation of DNA by RNA ligase
JOURNAL Patent: US 6368801-A 1 09-APR-2002;
FEATURES Location/Qualifiers
1.47
SOURCE /organism="unknown"

BASE COUNT 7 a 11 c 18 g 11 t
ORIGIN

Query Match 100.0%; Score 28; DB 6; Length 47;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGACCCCTATCAGGAGTACCA 28
DB 36 GCAGACCCCTATCAGGAGTACCA 9

RESULT 5
AX284180/c
LOCUS AX284180 47 bp DNA linear PAT 20-NOV-2001
DEFINITION Sequence 1 from Patent W00179420.
ACCESSION AX284180
VERSION AX284180.1 GI:17044868
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 Faruqi, A.F.
AUTHORS Detection and amplification of rna using target-mediated ligation of dna by rna ligase
JOURNAL Patent: WO 0179420-A 1 25-OCT-2001;
TITLE MOLECULAR STAGING, INC. (US)
FEATURES Location/Qualifiers
1.47
SOURCE /organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic Target"

BASE COUNT 7 a 11 c 18 g 11 t
ORIGIN

Query Match 100.0%; Score 28; DB 6; Length 47;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGACCCCTATCAGGAGTACCA 28
DB 36 GCAGACCCCTATCAGGAGTACCA 9

RESULT 6
I73305/c
LOCUS I73305 57 bp DNA linear PAT 03-APR-1998
DEFINITION Sequence 36 from patent US 5686272.

ACCESSION I73305
 VERSION I73305.1 GI:3009444
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 57)
 AUTHORS Marshall,R.L., Carrino,J.J. and Sustache,J.C.
 TITLE Amplification of RNA sequences using the ligase chain reaction
 JOURNAL Patent: US 5686272-A 36 11-NOV-1997;
 FEATURES
 source 1..57
 location/Qualifiers
 BASE COUNT 9 a 9 c 23 g 16 t
 ORIGIN
 Query Match 100.0%; Score 28; DB 6; Length 57;
 Best Local Similarity 100.0%; Pred. No. 0.11;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAAGCACCCTATCAGGAGTACCA 28
 ||||||||||||||||||||||||||||
 54 GCAAGCACCCTATCAGGAGTACCA 27
 RESULT 7
 LOCUS AX021668/c 75 bp DNA linear PAT 07-SEP-2000
 DEFINITION Sequence 47 from Patent WO9923250.
 ACCESSION AX021668
 VERSION AX021668.1 GI:10044951
 KEYWORDS
 SOURCE
 ORGANISM Hepatitis C virus.
 Hepatitis C virus.
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 1 (bases 1 to 75)
 Kessler,C., Bartl,K., Habershausen,G. and Orum,H.
 TITLE Specific and sensitive method for detecting nucleic acids
 JOURNAL Patent: WO 9923250-A 47 14-MAY-1999;
 KESSLER CHRISTOPH (DE); BARTL KNOT (DE); HABERSHAUSEN GERD (DE);
 ROCHE DIAGNOSTICS GMBH (DE); ORUM HENRIK (DK)
 FEATURES
 source 1..75
 location/Qualifiers
 BASE COUNT 13 a 20 c 25 g 17 t
 ORIGIN
 Query Match 100.0%; Score 28; DB 6; Length 75;
 Best Local Similarity 100.0%; Pred. No. 0.11;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAAGCACCCTATCAGGAGTACCA 28
 ||||||||||||||||||||||||||||
 30 GCAAGCACCCTATCAGGAGTACCA 3
 RESULT 8
 LOCUS E10302 186 bp RNA linear PAT 29-SEP-1997
 DEFINITION Anti-sense RNA against partial sequence of 5'UTR of hepatitis C
 virus.
 ACCESSION E10302
 VERSION E10302
 KEYWORDS JP 1995303485-A/3.
 SOURCE
 ORGANISM
 unclassified.
 1 (bases 1 to 186)
 Funahashi,S. and Hasegawa,A.
 TITLE HCV ANTI-SENSE RNA, MANIFESTATION VECTOR CONTAINING THE SAME, AND
 THERAPEUTIC AGENT FOR HCV-INVOLVED DISEASE CONTAINING THE RNA OR

JOURNAL
 Patent: JP 1995303485-A 3 21-NOV-1995;
 TONEN CORP
 COMMENT
 OS None
 OC Artificial sequences.
 PN JP 1995303485-A/3
 PD 21-NOV-1995
 PE 13-MAY-1994 JP 1994124609
 PI FUNAHASHI SHINICHI, HASEGAWA AKIRA
 PC C12N15/09,A61K31/70,A61K31/70,C07H21/04;
 CC strandedness: Single;
 CC topology: linear;
 CC hypothetical: No;
 CC anti-sense: Yes;
 FH Key
 FT source 1..186
 FT misc-feature 1..186
 FT /note="Anti-sense RNA against partial sequence
 of 5'UTR of
 hepatitis C virus".
 FT
 FEATURES
 source 1..186
 location/Qualifiers
 BASE COUNT 41 a 65 c 48 g 32 t
 ORIGIN
 Query Match 100.0%; Score 28; DB 6; Length 186;
 Best Local Similarity 100.0%; Pred. No. 0.096;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAAGCACCCTATCAGGAGTACCA 28
 ||||||||||||||||||||||||||||
 10 GCAAGCACCCTATCAGGAGTACCA 37
 RESULT 9
 LOCUS AX172757/c 187 bp mRNA linear PAT 03-JUL-2001
 DEFINITION Sequence 5' from Patent WO0144266.
 ACCESSION AX172757
 VERSION AX172757.1 GI:14597853
 KEYWORDS
 SOURCE
 ORGANISM
 synthetic construct.
 synthetic construct.
 artificial sequences.
 1 (bases 1 to 187)
 Karp,J.C. and Walker,S.C.
 TITLE Nucleic acid compounds and screening assays using the same
 JOURNAL Patent: WO 0144266-A 5 21-JUN-2001;
 Ribotargets Limited (GB)
 FEATURES
 source 1..187
 location/Qualifiers
 BASE COUNT 35 a 48 c 64 g 40 t
 ORIGIN
 Query Match 100.0%; Score 28; DB 6; Length 187;
 Best Local Similarity 100.0%; Pred. No. 0.096;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAAGCACCCTATCAGGAGTACCA 28
 ||||||||||||||||||||||||||||
 178 GCAAGCACCCTATCAGGAGTACCA 151
 RESULT 10
 LOCUS AF041266/c 191 bp ss-RNA linear VRL 06-DEC-2000
 AF041266

Query Match	100.0%	Score 28;	DB 14;	Length 191;					
Best Local Similarity	100.0%	Pred. No. 0.096;							
Matches	28;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;

Db 184 GCAAGCACCCTATCAGGCAGTACCACAA 15

RESULT 12
AE04127C/5

	760-137C	Hepatitis C virus isolate 760 5' untranslated region.	Linear VAL 08-DEC-2000
DEFINITION			
ACCESSION			

KEYWORDS
SOURCE
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839

Hepacivirus, no DNA stage; Flaviviridae,

Requeijo, M. P., Mando, O., Carballal, G. and Oubina, J. R. Espinola, L.,

J. Clin. Microbiol., 38 (12), 4560-4568 (2000)

2 (bases 1 to 191) .
Quarleri, T., Robertson R H and Cubie T D

Submitted (07-JAN-1998) Hepatitis Laboratory, Faculty of Medicine,
University of Buenos Aires, Paraguay 2155 11th Buenos Aires 1131

source
location/quarters
1. .191

```

/db_xref="taxon:11103"

```

ORIGIN

Best Local Similarity 100.0%; Pred. No. 0.096;

1 GCAAGCACCCTATCAGGCGAGTACCACAA 28

AF041277
OCUS
101 h- 22 mm

AF041277.1

ORGANISM Hepatitis C virus

REFERENCE
1 (bases 1 to 191)
hepacivirus.

TITLE Genomic and phylogenetic analysis of hepatitis C virus isolates

MEDLINE 20553364 (12/14) 4500-4500 (2000)

AUTHORS Quarleri, J., Robertson, B.H. and Oubina, J.R.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1998) Hepatitis Laboratory, Faculty of Medicine,
University of Buenos Aires, Paraguay 2155 11th, Buenos Aires 1121,
Argentina

FEATURES
source
1. .191
/organism="Hepatitis C virus"
/isolate="784"
/db_xref="taxon:11103"
1. .191

5'UTR
BASE COUNT 37 a 48 c 64 g 42 t
ORIGIN

Query Match 100.0%; Score 28; DB 14; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.096;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAAGCACCTATCGAGCAGTACCA 28
|||||
DB 184 GCAAGCACCTATCGAGCAGTACCA 157

RESULT 14
AF041279/c 191 bp ss-RNA linear VRL 06-DEC-2000
LOCUS AF041279
DEFINITION Hepatitis C virus isolate 784 5' untranslated region.
ACCESSION AF041279
VERSION AF041279.1 GI:2809138
KEYWORDS
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
VIRUSES: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE 1 (bases 1 to 191)
AUTHORS Quarleri, J.F., Robertson, B.H., Mathet, V.L., Feld, M., Espinola, L.,
Requejo, M.P., Mendo, O., Carballal, G. and Oubina, J.R.
TITLE Genomic and phylogenetic analysis of hepatitis C virus isolates
JOURNAL J. Clin. Microbiol. 38 (12), 4560-4568 (2000)
MEDLINE 20553364
PUBMED 11101596
REFERENCE 2 (bases 1 to 191)
AUTHORS Quarleri, J., Robertson, B.H. and Oubina, J.R.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1998) Hepatitis Laboratory, Faculty of Medicine,
University of Buenos Aires, Paraguay 2155 11th, Buenos Aires 1121,
Argentina

FEATURES
source
1. .191
/organism="Hepatitis C virus"
/isolate="784"
/db_xref="taxon:11103"
1. .191

5'UTR
BASE COUNT 34 a 51 c 64 g 42 t
ORIGIN

Query Match 100.0%; Score 28; DB 14; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.096;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAAGCACCTATCGAGCAGTACCA 28
|||||
DB 184 GCAAGCACCTATCGAGCAGTACCA 157

RESULT 15
AF041289/c 191 bp ss-RNA linear VRL 06-DEC-2000
LOCUS AF041289
DEFINITION Hepatitis C virus isolate 818 5' untranslated region.
ACCESSION AF041289
VERSION AF041289.1 GI:2809148
KEYWORDS

SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
VIRUSES: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE 1 (bases 1 to 191)
AUTHORS Quarleri, J.F., Robertson, B.H., Mathet, V.L., Feld, M., Espinola, L.,
Requejo, M.P., Mendo, O., Carballal, G. and Oubina, J.R.
TITLE Genomic and phylogenetic analysis of hepatitis C virus isolates
JOURNAL J. Clin. Microbiol. 38 (12), 4560-4568 (2000)
MEDLINE 20553364
PUBMED 11101596
REFERENCE 2 (bases 1 to 191)
AUTHORS Quarleri, J., Robertson, B.H. and Oubina, J.R.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1998) Hepatitis Laboratory, Faculty of Medicine,
University of Buenos Aires, Paraguay 2155 11th, Buenos Aires 1121,
Argentina

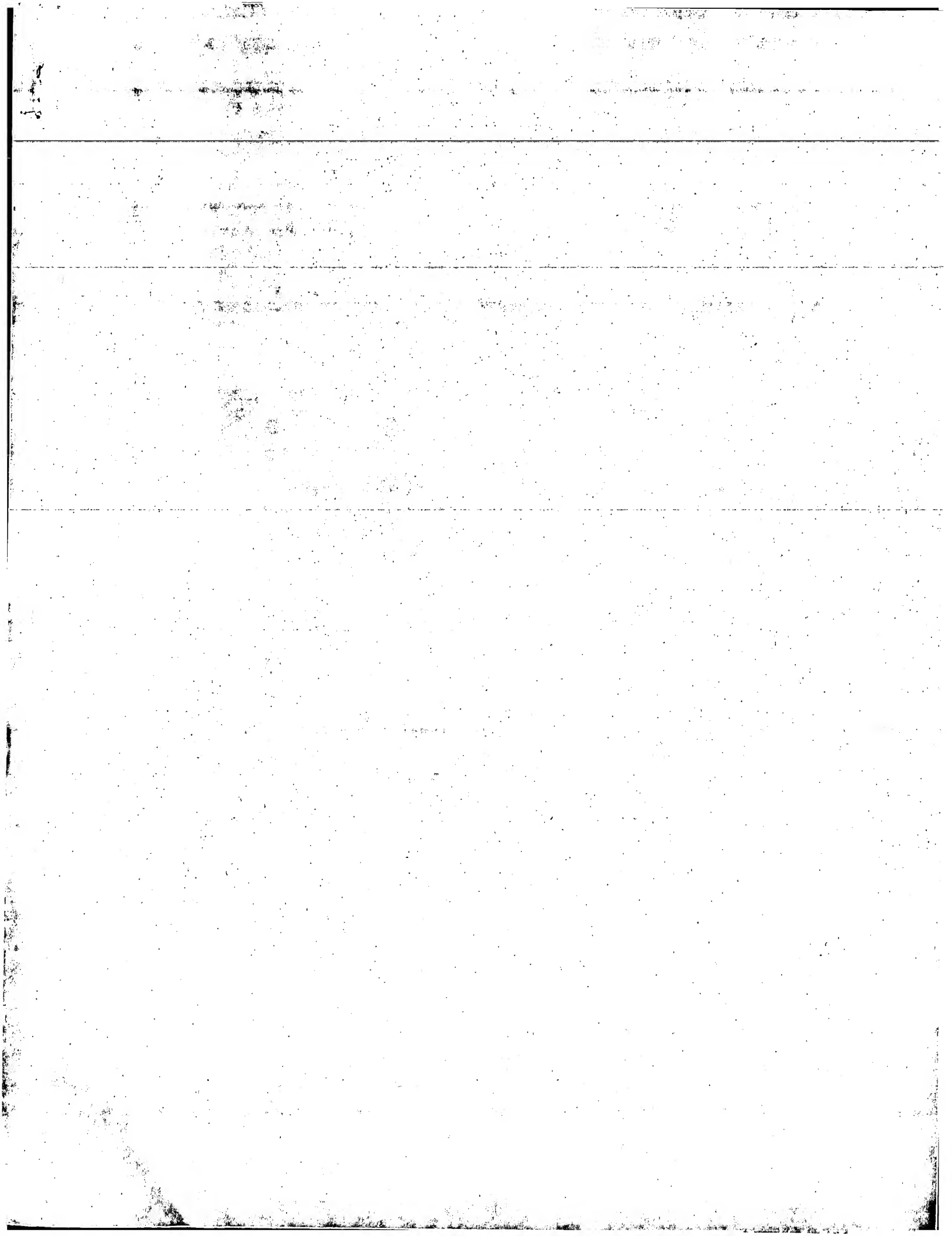
FEATURES
source
1. .191
/organism="Hepatitis C virus"
/isolate="818"
/db_xref="taxon:11103"
1. .191

5'UTR
BASE COUNT 34 a 51 c 64 g 42 t
ORIGIN

Query Match 100.0%; Score 28; DB 14; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.096;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAAGCACCTATCGAGCAGTACCA 28
|||||
DB 184 GCAAGCACCTATCGAGCAGTACCA 157

Search completed: July 10, 2003, 19:20:23
Job time : 243.625 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 225.294 Seconds

(without alignments)
3338.397 Million cell updates/sec

Title: US-10-087-631B-6

Perfect score: 26

Sequence: 1 gcagaagcgtctagccatgagcgtta 26

Scoring table: IDENTITY_NTC

Gapop 10.0, Gapect 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database :

GenEmbl:*

1: gb_da:*

2: gb_hlg:*

3: gb_in:*

4: gb_ov:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sy:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_mu:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pi:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hlg_hum:*

31: em_hlg_inv:*

32: em_hlg_inv:*

33: em_hlg_mus:*

34: em_hlg_pln:*

35: em_hlg_pln:*

36: em_hlg_rtd:*

37: em_hlg_rtd:*

38: em_hlg_vtl:*

39: em_hlg_vtl:*

40: em_hlg_hum:*

41: em_hlg_hum:*

42: em_hlg_hum:*

43: em_hlg_hum:*

44: em_hlg_hum:*

45: em_hlg_hum:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	26	100.0	26	6	AR054575	AR054575 Sequence
2	26	100.0	26	6	AR094137	AR094137 Sequence
3	26	100.0	26	6	AX147021	AX147021 Sequence
4	26	100.0	30	6	AX472307	AX472307 Sequence
5	26	100.0	51	6	AX021612	AX021612 Sequence
6	26	100.0	77	6	AX172761	AX172761 Sequence
7	26	100.0	123	14	HPCUT6CLN	M94468 Hepatitis C
8	26	100.0	137	14	HPCUT5CLN	M94467 Hepatitis C
9	26	100.0	142	14	S72378	S72378 (5' region)
10	26	100.0	155	6	AR095003	AR095003 Sequence
11	26	100.0	156	14	HPCCHAI1	M74253 Hepatitis C
12	26	100.0	160	14	HPCUT74CLN	M94466 Hepatitis C
13	26	100.0	171	14	HPCUT56CLN	M94461 Hepatitis C
14	26	100.0	176	14	HPCUT34CLN	M94462 Hepatitis C
15	26	100.0	176	14	HPCUT38CLN	M94499 Hepatitis C
16	26	100.0	189	14	AF077227	AF077227 Hepatitis C
17	26	100.0	189	14	AF021883	AF021883 Hepatitis C
18	26	100.0	190	14	AF021884	AF021884 Hepatitis C
19	26	100.0	190	14	AF021885	AF021885 Hepatitis C
20	26	100.0	190	14	AF021886	AF021886 Hepatitis C
21	26	100.0	190	14	AF021887	AF021887 Hepatitis C
22	26	100.0	190	14	AF021888	AF021888 Hepatitis C
23	26	100.0	190	14	AF021889	AF021889 Hepatitis C
24	26	100.0	190	14	AF021890	AF021890 Hepatitis C
25	26	100.0	190	14	AF021891	AF021891 Hepatitis C
26	26	100.0	190	14	AF021892	AF021892 Hepatitis C
27	26	100.0	190	14	AF021893	AF021893 Hepatitis C
28	26	100.0	190	14	AF021898	AF021898 Hepatitis C
29	26	100.0	190	14	AF021899	AF021899 Hepatitis C
30	26	100.0	190	14	AF021900	AF021900 Hepatitis C
31	26	100.0	190	14	AF021901	AF021901 Hepatitis C
32	26	100.0	190	14	AF021902	AF021902 Hepatitis C
33	26	100.0	190	14	AF021903	AF021903 Hepatitis C
34	26	100.0	190	14	AF021904	AF021904 Hepatitis C
35	26	100.0	194	6	AR066644	AR066644 Sequence
36	26	100.0	194	6	AR066645	AR066645 Sequence
37	26	100.0	194	6	AR066646	AR066646 Sequence
38	26	100.0	194	6	AR066647	AR066647 Sequence
39	26	100.0	194	6	AR066649	AR066649 Sequence
40	26	100.0	194	6	AR066650	AR066650 Sequence
41	26	100.0	194	6	AR066651	AR066651 Sequence
42	26	100.0	202	14	HPC5UTR1	D12506 Hepatitis C
43	26	100.0	202	14	HPC5UTR2	D12507 Hepatitis C
44	26	100.0	202	14	HPC5UTR2	D12507 Hepatitis C
45	26	100.0	202	14	HPC5UTR2	D12507 Hepatitis C

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
AR054575	AR054575	Sequence 1 from patent US 5837442.	AR054575	AR054575.1	GI:5980152	Unknown.	Unknown.	1 (bases 1 to 26)	Tsang,S.Yen.	Oligonucleotide primers for amplifying HCV nucleic acid	Patent: US 5837442-A 1 17-NOV-1998;	Location/Qualifiers
26 bp	DNA	Linear	PAT 29-SEP-1999									

Pred. No. is the number of results predicted by chance to have a

```

source
1. 26
BASE COUNT      /organism="unknown"
ORIGIN          7 a      6 c      8 g      5 t

Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 26;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 GCAGAAAGCGCTCTAGCCATGGCGTTA 26
1 GCAGAAAGCGCTCTAGCCATGGCGTTA 26

RESULT 2
AR094137
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. 26
BASE COUNT      /organism="unknown"
ORIGIN          7 a      6 c      8 g      5 t

Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 26;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 GCAGAAAGCGCTCTAGCCATGGCGTTA 26
1 GCAGAAAGCGCTCTAGCCATGGCGTTA 26

RESULT 3
AX147021
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. 26
BASE COUNT      /organism="synthetic construct"
ORIGIN          /db_xref="taxon:32630"
                  /note="Synthetic oligonucleotide primer (HCV forward)"
                  /note="derivatization with a p-(t-butyl)benzyl-residue"
                  /mod_base=OTHER
                  7 a      6 c      8 g      5 t

Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 26;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY
1 GCAGAAAGCGCTCTAGCCATGGCGTTA 26
1 GCAGAAAGCGCTCTAGCCATGGCGTTA 26

RESULT 4
AX472307
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. 30
BASE COUNT      /organism="Hepatitis C virus"
ORIGIN          /db_xref="taxon:11103"
                  8 a      7 c      9 g      6 t

Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 30;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 GCAGAAAGCGCTCTAGCCATGGCGTTA 26
3 GCAGAAAGCGCTCTAGCCATGGCGTTA 28

RESULT 5
AX021612
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. 51
BASE COUNT      /organism="Hepatitis C virus"
ORIGIN          /db_xref="taxon:11103"
                  11 a     12 c     15 g     13 t

Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 51;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 GCAGAAAGCGCTCTAGCCATGGCGTTA 26
8 GCAGAAAGCGCTCTAGCCATGGCGTTA 33

RESULT 6
AX172761

```

LOCUS AX172761 77 bp mRNA linear PAT 03-JUL-2001
 DEFINITION Sequence 9 from Patent WO0144266.
 ACCESSION AX172761
 VERSION AX172761.1 GI:14597857
 KEYWORDS
 ORGANISM
 SOURCE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 SOURCE
 1. 77
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="Probe"
 BASE COUNT 16 a 20 c 23 g 18 t
 ORIGIN
 Query Match 100.0%; Score 26; DB 6; Length 77;
 Best Local Similarity 100.0%; Pred. No. 0.056;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
 ||||||||||||||||||
 Db 26 GCAGAAAGCGTCTAGCCATGGCGTTA 51

RESULT 7
 LOCUS HPCUT6CLN 123 bp ss-RNA linear VRL 02-AUG-1993
 DEFINITION Hepatitis C virus (clone #6) nonstructural protein gene, 5' flank.
 ACCESSION M94468 M84479
 VERSION M94468.1 GI:329981
 KEYWORDS nonstructural protein.
 SOURCE Hepatitis C virus RNA.
 ORGANISM Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REMARK
 FEATURES
 SOURCE
 1. 123
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"
 BASE COUNT 28 a 35 c 36 g 24 t
 ORIGIN
 Query Match 100.0%; Score 26; DB 14; Length 123;
 Best Local Similarity 100.0%; Pred. No. 0.054;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
 ||||||||||||||||||
 Db 23 GCAGAAAGCGTCTAGCCATGGCGTTA 48

RESULT 8
 LOCUS HPCUT55CLN 137 bp ss-RNA linear VRL 02-AUG-1993
 DEFINITION Hepatitis C virus (clone #55) nonstructural protein gene, 5' flank.
 ACCESSION M94467 M84479
 VERSION M94467.1 GI:329979
 KEYWORDS nonstructural protein.

SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 FEATURES
 SOURCE
 1. 137
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"
 BASE COUNT 30 a 37 c 41 g 29 t
 ORIGIN
 Query Match 100.0%; Score 26; DB 14; Length 137;
 Best Local Similarity 100.0%; Pred. No. 0.054;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
 ||||||||||||||||||
 Db 23 GCAGAAAGCGTCTAGCCATGGCGTTA 48

RESULT 9
 LOCUS S72378 142 bp RNA linear VRL 24-JAN-1995
 DEFINITION [5' region] [hepatitis C virus HCV, host-human liver, Genomic RNA, 142 nt].
 ACCESSION S72378
 VERSION S72378.1 GI:632885
 KEYWORDS
 SOURCE
 ORGANISM
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepatitis C virus
 Hepacivirus.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REMARK
 FEATURES
 SOURCE
 1. 142
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"
 BASE COUNT 30 a 41 c 43 g 28 t
 ORIGIN
 Query Match 100.0%; Score 26; DB 14; Length 142;
 Best Local Similarity 100.0%; Pred. No. 0.054;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
 ||||||||||||||||||
 Db 4 GCAGAAAGCGTCTAGCCATGGCGTTA 29

RESULT 10
 LOCUS AR095003/C 155 bp DNA linear PAT 08-SEP-2000
 DEFINITION Sequence 41 from patent US 6001990.
 ACCESSION AR095003
 VERSION AR095003.1 GI:10022459

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
FEATURES
source

Unknown.
Unknown.
Unclassified.
1 (bases 1 to 155)
Wands, J.R., Wakita, T. and Moradpour, D.
Antisense inhibition of hepatitis C virus
Patent: US 6001990-A 41 14-DEC-1999;
Location/Qualifiers
1..155
/organism="unknown"

BASE COUNT
ORIGIN
29 a 43 c 51 g 32 t

Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 155;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY
1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
88 GCAGAAAGCGTCTAGCCATGGCGTTA 63

RESULT 11
HPCCH11
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

HPCCH11
Hepatitis C virus isolate 11 genome, 5' untranslated region.
M74253.1 GI:329742
Hepatitis C virus cDNA to genomic RNA.
Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C virus
1 (bases 1 to 156)
Chen, T.-A., Kolberg, J., Irvine, B., Stempien, M., Beall, E., Yano, M.,
Choo, Q.-L., Houghton, M., Kuo, G., Han, J.H. and Urdade, M.S.
Use of a signature nucleotide sequence of the hepatitis C virus for
the detection of viral RNA in human serum and plasma
Unpublished (1991)
Location/Qualifiers
1..156
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

BASE COUNT
ORIGIN
31 a 48 c 45 g 32 t

Query Match
Best Local Similarity 100.0%; Score 26; DB 14; Length 156;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY
1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
10 GCAGAAAGCGTCTAGCCATGGCGTTA 35

RESULT 12
HPCUT74CLN
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

HPCUT74CLN
Hepatitis C virus (clone #74) nonstructural protein gene, 5' flank.
M94466.1 GI:329982
nonstructural protein
Hepatitis C virus RNA.
Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C virus
1 (bases 1 to 160)
Martelli, M., Esteban, J.I., Quer, J., Genesca, J., Weiner, A.,
Esteban, R., Guardia, J. and Gomez, J.
Hepatitis C virus (HCV) circulates as a population of different but
closely related genomes: quasispecies nature of HCV genome
distribution

JOURNAL
MEDLINE
PubMed
FEATURES
source

J. Virol. 66 (5), 3225-3229 (1992)
92219420
1313927
Location/Qualifiers
1..160
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

BASE COUNT
ORIGIN
33 a 44 c 48 g 35 t

Query Match
Best Local Similarity 100.0%; Score 26; DB 14; Length 160;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY
1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
23 GCAGAAAGCGTCTAGCCATGGCGTTA 48

RESULT 13
HPCUT56CLN
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

HPCUT56CLN
Hepatitis C virus (clone #56) nonstructural protein gene, 5' flank.
M94464.1 GI:329980
nonstructural protein
Hepatitis C virus RNA.
Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C virus
1 (bases 1 to 171)
Martelli, M., Esteban, J.I., Quer, J., Genesca, J., Weiner, A.,
Esteban, R., Guardia, J. and Gomez, J.
Hepatitis C virus (HCV) circulates as a population of different but
closely related genomes: quasispecies nature of HCV genome
distribution
J. Virol. 66 (5), 3225-3229 (1992)
92219420
Location/Qualifiers
1..171
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

BASE COUNT
ORIGIN
37 a 51 c 47 g 36 t

Query Match
Best Local Similarity 100.0%; Score 26; DB 14; Length 171;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY
1 GCAGAAAGCGTCTAGCCATGGCGTTA 26
|||||
18 GCAGAAAGCGTCTAGCCATGGCGTTA 43

RESULT 14
HPCUT34CLN
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

HPCUT34CLN
Hepatitis C virus (clone #34) nonstructural protein gene, 5' flank.
M94461.1 GI:329976
nonstructural protein
Hepatitis C virus RNA.
Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C virus
1 (bases 1 to 176)
Martelli, M., Esteban, J.I., Quer, J., Genesca, J., Weiner, A.,
Esteban, R., Guardia, J. and Gomez, J.
Hepatitis C virus (HCV) circulates as a population of different but
closely related genomes: quasispecies nature of HCV genome
distribution
J. Virol. 66 (5), 3225-3229 (1992)

MEDLINE 92219420
 PUBMED 1313927
 FEATURES
 source Location/Qualifiers
 1.176

BASE COUNT 37 a 50 c 51 g 38 t
 ORIGIN
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"

Query Match 100.0%; Score 26; DB 14; Length 176;
 Best Local Similarity 100.0%; Pred. No. 0.053;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCCATGCGCTTA 26
 |||||
 Db 23 GCAGAAAGCGCTAGCCATGCGCTTA 48

RESULT 15
 HPCUT38CLN 176 bp ss-RNA linear VR1_02-AUG-1993
 LOCUS Hepatitis C virus (clone #38) nonstructural protein gene, 5' flank.
 DEFINITION M94462 M84479
 ACCESSION M94462.1 GI:329977
 VERSION nonstructural protein.
 KEYWORDS Hepatitis C virus RNA.
 SOURCE Hepatitis C virus
 ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepcivirus.
 1 (bases 1 to 176)

REFERENCE
 AUTHORS Martell,M., Esteban,J.I., Quer,J., Genesca,J., Weiner,A.,
 Esteban,R., Guardia,J. and Gomez,J.
 TITLE Hepatitis C virus (HCV) circulates as a population of different but
 closely related genomes: quasispecies nature of HCV genome
 distribution

JOURNAL J. Virol. 66 (5), 3225-3229 (1992)
 MEDLINE 92219420
 PUBMED 1313927

FEATURES
 source Location/Qualifiers
 1.176

BASE COUNT 36 a 51 c 51 g 38 t
 ORIGIN
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"

Query Match 100.0%; Score 26; DB 14; Length 176;
 Best Local Similarity 100.0%; Pred. No. 0.053;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGCTAGCCATGCGCTTA 26
 |||||
 Db 23 GCAGAAAGCGCTAGCCATGCGCTTA 48

Search completed: July 10, 2003, 19:20:22.
 Job time : 227.294 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 ; Search time 107.773 Seconds
(without alignments)
685.785 Million cell updates/sec

Title: US-10-087-631b-10

Perfect score: 241
Sequence: 1 gcagaagcgtctagcctg.....ctgcctgtaggtgcttgc 241

Scoring table: IDENTITY_NUC
Gapop 10.0 / Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCUTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/Backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	241	100.0	324	2	US-08-470-426B-15
2	241	100.0	341	3	US-08-854-531-4
3	241	100.0	341	3	US-09-014-416-47
4	241	100.0	341	3	US-09-014-416-48
5	241	100.0	341	4	US-08-869-380-4
6	241	100.0	341	5	PCT-US95-13552-4
7	241	100.0	342	3	US-08-474-700B-39
8	241	100.0	686	4	US-08-988-321B-37
9	241	100.0	686	4	US-08-397-220B-25
10	241	100.0	686	4	US-08-650-093C-25
11	241	100.0	780	3	US-08-474-700B-45
12	241	100.0	923	4	US-08-869-380-1
13	241	100.0	923	4	PCT-US95-13552-14
14	241	100.0	1499	1	US-08-324-977-3
15	241	100.0	1499	2	US-08-384-616-3
16	241	100.0	1499	2	US-08-904-686A-3
17	241	100.0	1499	4	US-09-315-850-3
18	241	100.0	1863	2	US-08-470-426B-14
19	241	100.0	9416	1	US-08-324-977-1
20	241	100.0	9416	2	US-08-384-616-1
21	241	100.0	9416	2	US-08-904-686A-1
22	241	100.0	9416	4	US-09-315-850-1
23	241	100.0	9416	4	US-08-823-885A-27
24	241	100.0	9595	3	US-09-014-416-4
25	241	100.0	9599	3	US-09-014-416-6
26	239.4	99.3	244	4	US-09-034-205-26
27	239.4	99.3	244	4	US-09-034-205-27

28	239.4	99.3	244	4	US-08-934-097A-26	Sequence 26, Appl
29	239.4	99.3	244	4	US-08-934-097A-27	Sequence 27, Appl
30	239.4	99.3	244	4	US-08-851-568-26	Sequence 26, Appl
31	239.4	99.3	244	4	US-08-851-568-27	Sequence 27, Appl
32	239.4	99.3	244	4	US-09-677-218B-26	Sequence 26, Appl
33	239.4	99.3	244	4	US-09-677-218B-27	Sequence 27, Appl
34	239.4	99.3	244	4	US-09-677-192-26	Sequence 26, Appl
35	239.4	99.3	244	4	US-09-677-192-27	Sequence 27, Appl
36	239.4	99.3	281	2	US-08-757-653-123	Sequence 123, App
37	239.4	99.3	281	2	US-08-757-653-128	Sequence 128, App
38	239.4	99.3	281	2	US-08-757-653-129	Sequence 129, App
39	239.4	99.3	281	4	US-08-520-946-123	Sequence 123, App
40	239.4	99.3	281	4	US-08-520-946-128	Sequence 128, App
41	239.4	99.3	281	4	US-08-520-946-129	Sequence 129, App
42	239.4	99.3	286	4	US-09-034-205-21	Sequence 21, Appl
43	239.4	99.3	286	4	US-08-934-097A-21	Sequence 21, Appl
44	239.4	99.3	286	4	US-08-851-568-21	Sequence 21, Appl
45	239.4	99.3	286	4	US-09-677-218B-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-470-426B-15
Sequence 15, Application US/08470426B
Patent No. 5856458
GENERAL INFORMATION:
APPLICANT: Okamoto, Hiroaki
APPLICANT: Nakamura, Tetsuo
TITLE OF INVENTION: OLIGONUCLEOTIDE PRIMERS, AND THEIR
TITLE OF INVENTION: APPLICATION FOR HIGH-FIDELITY DETECTION OF NON-A, NON-B
TITLE OF INVENTION: HEPATITIS VIRUS
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Beveridge, Degrandi, Wellacher & Young,
ADDRESS: L.L.P.
STREET: 1650 M Street, N.W., Suite 800
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,426B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-153402
FILING DATE: 12-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Wellacher, Robert G.
REGISTRATION NUMBER: 20,531
REFERENCE/DOCKET NUMBER: 06/59-47083.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2811
TELEFAX: (202) 659-1462
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-470-426B-15
Query Match 100.0%; Score 241; DB 2; Length 324;
Best Local Similarity 100.0%; Pred. No. 1.5e-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGGTTAGTATGAGTGTGCTGACCTCCAGAGACCCCTT 60
DB 51 GCAGAAAGCGTCTAGCCATGCGGTTAGTATGAGTGTGCTGACCTCCAGAGACCCCTT 110
QY 61 CCCGGAGAGCCATAGTGTGCTGCGGAACCGGTAGTACACCGGAATTTGCCAGAGAC 120
DB 111 CCCGGAGAGCCATAGTGTGCTGCGGAACCGGTAGTACACCGGAATTTGCCAGAGAC 170
QY 121 GGGTCCTTTCTTGATACACCGGTCAATGCTGAGATTTGGCGCTCCCGGAGAC 180
DB 171 GGGTCCTTTCTTGATACACCGGTCAATGCTGAGATTTGGCGCTCCCGGAGAC 230
QY 181 TGTAGCCGAGTATGTTGGTCCGGAAGCGCTTGCTGCTGCTGATAGGCTGCTT 240
DB 231 TGTAGCCGAGTATGTTGGTCCGGAAGCGCTTGCTGCTGCTGATAGGCTGCTT 290
QY 241 C 241
DB 291 C 291

RESULT 2

US-08-854-531-4
Sequence 4, Application US/08854531
Patent No. 6025341

GENERAL INFORMATION:

APPLICANT: Wands, Jack
APPLICANT: Tokushige, Katsutoshi
TITLE OF INVENTION: CHIMERIC HEPATITIS B/HEPATITIS C VIRUS VACCINE
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6025341
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/854,531
FILING DATE: Herewith
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: APOL-0214
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 341 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-854-531-4

Query Match 100.0% Score 241; DB 3; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.6e-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGGTTAGTATGAGTGTGCTGACCTCCAGAGACCCCTT 60
DB 68 GCAGAAAGCGTCTAGCCATGCGGTTAGTATGAGTGTGCTGACCTCCAGAGACCCCTT 127
QY 61 CCCGGAGAGCCATAGTGTGCTGCGGAACCGGTAGTACACCGGAATTTGCCAGAGAC 120

DB 128 CCCGGAGAGCCATAGTGTGCTGCGGAACCGGTAGTACACCGGAATTTGCCAGAGAC 187
QY 121 GGGTCCTTTCTTGATACACCGGTCAATGCTGAGATTTGGCGCTCCCGGAGAC 180
DB 188 GGGTCCTTTCTTGATACACCGGTCAATGCTGAGATTTGGCGCTCCCGGAGAC 247
QY 181 TGTAGCCGAGTATGTTGGTCCGGAAGCGCTTGCTGCTGCTGATAGGCTGCTT 240
DB 248 TGTAGCCGAGTATGTTGGTCCGGAAGCGCTTGCTGCTGCTGATAGGCTGCTT 307
QY 241 C 241
DB 308 C 308

RESULT 3

US-09-014-416-47
Sequence 47, Application US/09014416
Patent No. 6153421

GENERAL INFORMATION:

APPLICANT: Yanagii, Masayuki
APPLICANT: Bukh, Jens
APPLICANT: Emerson, Susanne U.
APPLICANT: Purcell, Robert H.
TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
FILE REFERENCE: 20264276
CURRENT APPLICATION NUMBER: US/09/014,416
EARLIER FILING DATE: 1998-01-27
EARLIER APPLICATION NUMBER: US 60/053,062
NUMBER OF SEQ. ID NOS: 65
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 47
LENGTH: 341
TYPE: DNA
ORGANISM: Hepatitis C virus
US-09-014-416-47

Query Match 100.0% Score 241; DB 3; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.6e-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGGTTAGTATGAGTGTGCTGACCTCCAGAGACCCCTT 60
DB 68 GCAGAAAGCGTCTAGCCATGCGGTTAGTATGAGTGTGCTGACCTCCAGAGACCCCTT 127
QY 61 CCCGGAGAGCCATAGTGTGCTGCGGAACCGGTAGTACACCGGAATTTGCCAGAGAC 120
DB 128 CCCGGAGAGCCATAGTGTGCTGCGGAACCGGTAGTACACCGGAATTTGCCAGAGAC 187
QY 121 GGGTCCTTTCTTGATACACCGGTCAATGCTGAGATTTGGCGCTCCCGGAGAC 180
DB 188 GGGTCCTTTCTTGATACACCGGTCAATGCTGAGATTTGGCGCTCCCGGAGAC 247
QY 181 TGTAGCCGAGTATGTTGGTCCGGAAGCGCTTGCTGCTGCTGATAGGCTGCTT 240
DB 248 TGTAGCCGAGTATGTTGGTCCGGAAGCGCTTGCTGCTGCTGATAGGCTGCTT 307
QY 241 C 241
DB 308 C 308

RESULT 4

US-09-014-416-48
Sequence 48, Application US/09014416
Patent No. 6153421

GENERAL INFORMATION:

APPLICANT: Yanagii, Masayuki
APPLICANT: Bukh, Jens
APPLICANT: Emerson, Susanne U.
APPLICANT: Purcell, Robert H.

TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
FILE OF INVENTION: USES THEREOF
FILE REFERENCE: 20264276
CURRENT APPLICATION NUMBER: US/09/014,416
CURRENT FILING DATE: 1998-01-27
EARLIER APPLICATION NUMBER: US 60/053,062
EARLIER FILING DATE: 1997-07-18
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 48
LENGTH: 341
TYPE: DNA
ORGANISM: Hepatitis C virus
US-09-014-416-48

Query Match 100.0%; Score 241; DB 3; Length 341;
Best Local Similarity 100.0%; Pred. No. 1,66-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAACGCTCTACCATGCGCTTATGATGATGCTGACGCTCCAGGACCCCT 60
DB 68 GCAGAAACGCTCTACCATGCGCTTATGATGATGCTGACGCTCCAGGACCCCT 127
QY 61 CCCGGAGAGCCATAGTGTCTGCGAACCGGTGAGTACACCGGAATTGCCAGAGAC 120
DB 128 CCCGGAGAGCCATAGTGTCTGCGAACCGGTGAGTACACCGGAATTGCCAGAGAC 187
QY 121 GGGTCTCTTCTTGATCAACCCGCTCAATGCCGTGAGATTGGCGCTGCCCGGAGAC 180
DB 188 GGGTCTCTTCTTGATCAACCCGCTCAATGCCGTGAGATTGGCGCTGCCCGGAGAC 247
QY 181 TGCTAGCCGAGTAGTGTGGTGGCGAAAGCGCTGTGCTACTGCTGATAGGGTCTTG 240
DB 248 TGCTAGCCGAGTAGTGTGGTGGCGAAAGCGCTGTGCTACTGCTGATAGGGTCTTG 307
QY 241 C 241
DB 308 C 308

RESULT 5

US-08-869-380-4
Sequence 4, Application US/08869380
Patent No. 6235888
GENERAL INFORMATION:
APPLICANT: Pachuk, Catherine J.
APPLICANT: Wands, Jack
APPLICANT: Wakita, Takaji
APPLICANT: Zurawski, Jr., Vincent R.
APPLICANT: Coney, Leslie R.
TITLE OF INVENTION: Hepatitis C Virus Vaccine
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
ADDRESS: No. 6235888
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/869,380
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/318,248
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: APOL-0189
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 341 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-869-380-4

Query Match 100.0%; Score 241; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 1,66-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAACGCTCTACCATGCGCTTATGATGATGCTGACGCTCCAGGACCCCT 60
DB 68 GCAGAAACGCTCTACCATGCGCTTATGATGATGCTGACGCTCCAGGACCCCT 127
QY 61 CCCGGAGAGCCATAGTGTCTGCGAACCGGTGAGTACACCGGAATTGCCAGAGAC 120
DB 128 CCCGGAGAGCCATAGTGTCTGCGAACCGGTGAGTACACCGGAATTGCCAGAGAC 187
QY 121 GGGTCTCTTCTTGATCAACCCGCTCAATGCCGTGAGATTGGCGCTGCCCGGAGAC 180
DB 188 GGGTCTCTTCTTGATCAACCCGCTCAATGCCGTGAGATTGGCGCTGCCCGGAGAC 247
QY 181 TGCTAGCCGAGTAGTGTGGTGGCGAAAGCGCTGTGCTACTGCTGATAGGGTCTTG 240
DB 248 TGCTAGCCGAGTAGTGTGGTGGCGAAAGCGCTGTGCTACTGCTGATAGGGTCTTG 307
QY 241 C 241
DB 308 C 308

RESULT 6

PCT-US95-13552-4
Sequence 4, Application PC/TUS9513552
GENERAL INFORMATION:
APPLICANT: Wands, Jack
APPLICANT: Tokushige, Katsutoshi
APPLICANT: Wakita, Takaji
APPLICANT: Pachuk, Catherine J.
APPLICANT: Zurawski, Jr., Vincent R.
APPLICANT: Coney, Leslie R.
TITLE OF INVENTION: HEPATITIS VIRUS VACCINES
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
ADDRESS: Norris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13552
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/318,248
FILING DATE: 05-OCT-1994
CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/467,859
 FILING DATE: 06-JUN-1995
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Deluca, Mark
 REGISTRATION NUMBER: 33,229
 REFERENCE/DOCKET NUMBER: APOI-0238
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 341 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 PCT-US95-13552-4

Query Match 100.0%; Score 241; DB 5; Length 341;
 Best Local Similarity 100.0%; Pred. No. 1.6e-68;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCATGCGGTTAGTGTGAGTGTGCGAGCCTCCAGAGACCCCTT 60
 DB 68 GCAGAAAGCGTCTAGCATGCGGTTAGTGTGAGTGTGCGAGCCTCCAGAGACCCCTT 127
 OY 61 CCCGGAGAGCCATAGTGTGCGGAAACCGGTGAGTACACCGGAATTGCCAGAGACGCC 120
 DB 128 CCCGGAGAGCCATAGTGTGCGGAAACCGGTGAGTACACCGGAATTGCCAGAGACGCC 187
 OY 121 GGGTCCTTCTTGATCAACCCGCTCATGCTGAGATTTGGGCGTCCCGCGAGAC 180
 DB 188 GGGTCCTTCTTGATCAACCCGCTCATGCTGAGATTTGGGCGTCCCGCGAGAC 247
 OY 181 TGTACCGAGTAGTGTGCGGAAACCGGTGAGTACACCGGAATTGCCAGAGACGCC 240
 DB 248 TGTACCGAGTAGTGTGCGGAAACCGGTGAGTACACCGGAATTGCCAGAGACGCC 307
 OY 241 C 241
 DB 308 C 308

RESULT 7

US-08-474-700B-39/C
 Sequence 39, Application US/08474700B
 Patent No. 6001990

GENERAL INFORMATION:

APPLICANT: Hands, Jack
 APPLICANT: Moradpour, Darius
 TITLE OF INVENTION: ANTISENSE INHIBITION OF HEPATITIS C
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 MB
 COMPUTER: IBM PS/2 Model 502 or 555X
 OPERATING SYSTEM: MS-DOS (Version 5.0)
 SOFTWARE: WordPerfect (Version 5.1)
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/474,700B
 FILING DATE: 07-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/240,382
 FILING DATE: 10 May 1994

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.
 REGISTRATION NUMBER: 34,819
 REFERENCE/DOCKET NUMBER: 00786/279001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 542-5070
 TELEFAX: (617) 542-8906
 TELEEX: 200154
 INFORMATION FOR SEQ ID NO: 39:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 342 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 US-08-474-700B-39

Query Match 100.0%; Score 241; DB 3; Length 342;
 Best Local Similarity 100.0%; Pred. No. 1.6e-68;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCAGAAAGCGTCTAGCATGCGGTTAGTGTGAGTGTGCGAGCCTCCAGAGACCCCTT 60
 DB 275 GCAGAAAGCGTCTAGCATGCGGTTAGTGTGAGTGTGCGAGCCTCCAGAGACCCCTT 216
 OY 61 CCCGGAGAGCCATAGTGTGCGGAAACCGGTGAGTACACCGGAATTGCCAGAGACGCC 120
 DB 215 CCCGGAGAGCCATAGTGTGCGGAAACCGGTGAGTACACCGGAATTGCCAGAGACGCC 156
 OY 121 GGGTCCTTCTTGATCAACCCGCTCATGCTGAGATTTGGGCGTCCCGCGAGAC 180
 DB 155 GGGTCCTTCTTGATCAACCCGCTCATGCTGAGATTTGGGCGTCCCGCGAGAC 96
 OY 181 TGTACCGAGTAGTGTGCGGAAACCGGTGAGTACACCGGAATTGCCAGAGACGCC 240
 DB 95 TGTACCGAGTAGTGTGCGGAAACCGGTGAGTACACCGGAATTGCCAGAGACGCC 36
 OY 241 C 241
 DB 35 C 35

RESULT 8

US-08-988-321B-37
 Sequence 37, Application US/08988321B
 Patent No. 6174868

GENERAL INFORMATION:

APPLICANT: Kevin P. Anderson et al.
 TITLE OF INVENTION: Compositions And Methods For Treatment Of Hepatitis C V
 NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Law Offices of Jane Massey Licata
 STREET: 66 East Main Street
 CITY: Marlton
 STATE: NJ
 COUNTRY: USA
 ZIP: 08053

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
 COMPUTER: IBM COMPATIBLE
 OPERATING SYSTEM: WINDOWS 95
 SOFTWARE: WORDPERFECT 6.1 FOR WINDOWS
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/988,321B
 FILING DATE: December 10, 1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/650,093
 FILING DATE: May 17, 1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/452,841
 FILING DATE: May 30, 1995
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/397,220
FILING DATE: March 9, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/945,289
FILING DATE: September 10, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: ISPH-0245
TELECOMMUNICATION INFORMATION:
TELEPHONE: (856) 810-1515
TELEFAX: (856) 810-1454
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 686
TYPE: nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: No
US-08-988-321B-37

Query Match 100.0%; Score 241; DB 4; Length 686;
Best Local Similarity 79.3%; Pred. No. 1.9e-68;
Matches 191; Conservative 50; Mismatches 0; Indels 0; Gaps 0;

1 GCAGAAAGCGTCTAGCCATGCGCTAGTATGAGTGTGCGAGCCCTCGAGACCCCT 60
68 GCAGAAAGCGTCTAGCCATGCGCTAGTATGAGTGTGCGAGCCCTCGAGACCCCT 127
61 CCCGGAGAGCCATAGTGTGCGAGACCCGCTGAGTACACCGGATTCGAGAGAC 120
128 CCCGGAGAGCCATAGTGTGCGAGACCCGCTGAGTACACCGGATTCGAGAGAC 187
121 GGGTCTTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCTCCCCCGAGAC 180
188 GGGTCTTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCTCCCCCGAGAC 247
181 TGCTAGCCGAGTGTGAGTGTGCGAGAAAGCGCTGTGCTAGTGTGAGTGTG 240
248 UGCUAGCCGAGUAGUGUGGUGCGGAAAGGCCUUGGUGUACUUGCCUAGUGUG 307
241 C 241
308 C 308

RESULT 9

US-08-397-220B-25
Sequence 25, Application US/08397220B
Patent No. 6284458

GENERAL INFORMATION:

APPLICANT: Anderson et al.

TITLE OF INVENTION: Compositions And Methods For Treatment
Of Hepatitis C Virus-Associated Diseases

NUMBER OF SEQUENCES: 98

CORRESPONDENCE ADDRESS:

ADDRESSEE: Jane Massey Licata, Esq.

STREET: 210 Lake Drive East, Suite 201

CITY: Cherry Hill

STATE: NJ

COUNTRY: USA

ZIP: 08002

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE

COMPUTER: IBM 486

OPERATING SYSTEM: WINDOWS FOR WORKGROUPS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/397,220B

FILING DATE: 09-Mar-1995

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP93/01293

FILING DATE: 10-Sep-93
APPLICATION NUMBER: JP 5-87195
FILING DATE: 14-Apr-93
APPLICATION NUMBER: 07/945,289
FILING DATE: 10-Sep-92
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: ISPH-0031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 686
TYPE: nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: No
SEQUENCE DESCRIPTION: SEQ ID NO: 25:
US-08-397-220B-25

Query Match 100.0%; Score 241; DB 4; Length 686;
Best Local Similarity 79.3%; Pred. No. 1.9e-68;
Matches 191; Conservative 50; Mismatches 0; Indels 0; Gaps 0;

1 GCAGAAAGCGTCTAGCCATGCGCTAGTATGAGTGTGCGAGCCCTCGAGACCCCT 60
68 GCAGAAAGCGTCTAGCCATGCGCTAGTATGAGTGTGCGAGCCCTCGAGACCCCT 127
61 CCCGGAGAGCCATAGTGTGCGAGACCCGCTGAGTACACCGGATTCGAGAGAC 120
128 CCCGGAGAGCCATAGTGTGCGAGACCCGCTGAGTACACCGGATTCGAGAGAC 187
121 GGGTCTTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCTCCCCCGAGAC 180
188 GGGTCTTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGCTCCCCCGAGAC 247
181 TGCTAGCCGAGTGTGAGTGTGCGAGAAAGCGCTGTGCTAGTGTGAGTGTG 240
248 UGCUAGCCGAGUAGUGUGGUGCGGAAAGGCCUUGGUGUACUUGCCUAGUGUG 307
241 C 241
308 C 308

RESULT 10

US-08-650-093C-25
Sequence 25, Application US/08650093C
Patent No. 6391542

GENERAL INFORMATION:

APPLICANT: Kevin P. Anderson et al.

TITLE OF INVENTION: Compositions And Methods For Treatment Of
Hepatitis C Virus-Associated Diseases

NUMBER OF SEQUENCES: 118

CORRESPONDENCE ADDRESS:

ADDRESSEE: LICATA & TYRRELL P.C.

STREET: 66 E. Main Street

CITY: Marlton

STATE: NJ

COUNTRY: USA

ZIP: 08053

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: WORDPERFECT 6.1 for Windows

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/650,093C

FILING DATE: 17-May-1996

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

us-10-087-631b-10.rni

Page 6

Best match	100.00;	Score 241;	DB 4;	Length 686;
Best Local	79.38;	Pred. No. 1.9e-68;		
Matches 191;	Conservative 50;	Mismatches 0;		

[illegible]

RESULT 11
 US-08-474-700B-45.
 Sequence 45, Application US/08474700B
 Patent No. 6001990
 GENERAL INFORMATION:
 APPLICANT: Wands, Jack
 APPLICANT: Wakita, Takaji
 APPLICANT: Moradpour, Darius
 TITLE OF INVENTION: ANTISENSE INHIBITION OF HEPATITIS C
 TITLE OF INVENTION: VIRUS
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 MB
 COMPUTER: IBM PS/2 Model 502 or 555X
 OPERATING SYSTEM: MS-DOS (Version 5.0)
 SOFTWARE: Wordperfect (Version 5.1)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/474,700B

Query Match	100.0%	Score 241;	DB 3;	Length 780;
Best Local Similarity	100.0%	Pred. No. 2e-68;		
Matches 241;	Conservative	0;	Mismatches	0;

OY	1	GGAGAAAGGTTACCAATGGCGTTAGATGATGTCTGCAGCCCTCCAGAACCCCCCT	60
Db	68	GGAGAAACGCTCTAGCCCATGGCGTTAGATGATGTCTGCAGCCCTCCAGAACCCCCCT	127
OY	61	CCCGGGAAGCATATGTTGGTCTGCCGAACCGGATGATACACCGGAATTGCCAGAGACAC	120
Db	128	CCCGGGAAGCATATGTTGGTCTGCCGAACCGGATGATACACCGGAATTGCCAGAGACAC	187
OY	121	GGGTCTCTTTGGATCAACCCGCTCAATGCTCTGGAGATTGGGCGTGGCCCGCGAGAC	180
Db	188	GGGTCTCTTTGGATCAACCCGCTCAATGCTCTGGAGATTGGGCGTGGCCCGCGAGAC	247
OY	181	TGCTAGCCGATAGTGTGGTGCAGAAAGCCCTTGTGGATACCCGATAGGGTGGCTTG	240
Db	248	TGCTAGCCGATAGTGTGGTGCAGAAAGCCCTTGTGGATACCCGATAGGGTGGCTTG	307
OY	241	C 241	
Db	308	C 308	

RESULT- 12-
 US-08-869-380-1
 Sequence 1, Application US/0869380
 Patent No. 6235888
 GENERAL INFORMATION:
 APPLICANT: Pachut, Catherine J.
 APPLICANT: Mands, Jack
 APPLICANT: Wakita, Takaji
 APPLICANT: Zurawski, J., Vincent R.
 APPLICANT: Coney, Leslie R.
 TITLE OF INVENTION: Hepatitis C Virus Vaccine
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
 ADDRESSEE: No. 6235888ris
 STREET: One Liberty Place, 46th floor
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/869,380
 FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/318,248
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: APOL-0189
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 923 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 342..914
US-08-869-380-1

Query Match 100.0%; Score 241; DB 4; Length 923;
Best Local Similarity 100.0%; Pred. No. 2,1e-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTGACCATGCGCTTATGATAGTGTGCTGACGCTCCAGACCCCCCT 60
DB 68 GCAGAAAGCGCTGACCATGCGCTTATGATAGTGTGCTGACGCTCCAGACCCCCCT 127
QY 61 CCGGGAGAGCCATAGTGTGCTGCGGAACCGGTGATGACCGGATGCCAGAGACGCC 120
DB 128 CCGGGAGAGCCATAGTGTGCTGCGGAACCGGTGATGACCGGATGCCAGAGACGCC 187
QY 121 GGGTCCTTTCTTGATCAACCCGCTCATGCTGAGATTGGGCGTCCCGCGAGAC 180
DB 188 GGGTCCTTTCTTGATCAACCCGCTCATGCTGAGATTGGGCGTCCCGCGAGAC 247
QY 181 TGTAGCCGAGTAGTGTGCTGCGGAACCGGTGATGACCGGATGCCAGAGACGCC 240
DB 248 TGTAGCCGAGTAGTGTGCTGCGGAACCGGTGATGACCGGATGCCAGAGACGCC 307
QY 241 C 241
DB 308 C 308

RESULT 13
PCT-US95-13552-14
Sequence 14, Application PC/TUS9513552
GENERAL INFORMATION:
APPLICANT: Wands, Jack
APPLICANT: Tokushige, Katsutoshi
APPLICANT: Wakita, Takaji
APPLICANT: Pachuk, Catherine J.
APPLICANT: Zurawski, Jr., Vincent R.
TITLE OF INVENTION: HEPATITIS VIRUS VACCINES
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13552
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/318,248
FILING DATE: 05-OCT-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/467,859
FILING DATE: 06-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: APOL-0238
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 923 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 342..914
PCT-US95-13552-14

Query Match 100.0%; Score 241; DB 5; Length 923;
Best Local Similarity 100.0%; Pred. No. 2,1e-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTGACCATGCGCTTATGATAGTGTGCTGACGCTCCAGACCCCCCT 60
DB 68 GCAGAAAGCGCTGACCATGCGCTTATGATAGTGTGCTGACGCTCCAGACCCCCCT 127
QY 61 CCGGGAGAGCCATAGTGTGCTGCGGAACCGGTGATGACCGGATGCCAGAGACGCC 120
DB 128 CCGGGAGAGCCATAGTGTGCTGCGGAACCGGTGATGACCGGATGCCAGAGACGCC 187
QY 121 GGGTCCTTTCTTGATCAACCCGCTCATGCTGAGATTGGGCGTCCCGCGAGAC 180
DB 188 GGGTCCTTTCTTGATCAACCCGCTCATGCTGAGATTGGGCGTCCCGCGAGAC 247
QY 181 TGTAGCCGAGTAGTGTGCTGCGGAACCGGTGATGACCGGATGCCAGAGACGCC 240
DB 248 TGTAGCCGAGTAGTGTGCTGCGGAACCGGTGATGACCGGATGCCAGAGACGCC 307
QY 241 C 241
DB 308 C 308

RESULT 14
US-08-324-977-3
Sequence 3, Application US/08324977
Patent No. 5747339
GENERAL INFORMATION:
APPLICANT: OKAYAMA, HIROTO
APPLICANT: FUKE, ISAO
APPLICANT: MORI, CHISATO
APPLICANT: TAKAMIZAWA, AKAHISA
APPLICANT: YOSHIDA, IWAO
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Armstrong, Westerman, Hattori, McIreland &
STREET: 1725 K St. N.W. Suite 1000

CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/324,977
FILING DATE: 18-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-167466
FILING DATE: 25-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-230921
FILING DATE: 31-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-305605
FILING DATE: 09-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/099,706
FILING DATE: 30-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/769,996
FILING DATE: 02-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/635,451
FILING DATE: 28-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Stevens-Smith, Theresa M.
REGISTRATION NUMBER: 36,281
REFERENCE/DOCKET NUMBER: 900703D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
TELEX: 440142
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1499 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA from genomic RNA
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1499
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..1499
OTHER INFORMATION: /note= "sequence = 1 - 1499 of SEQ
US-08-324-977-3
Query Match 100.0% Score 241; DB 1; Length 1499;
Best Local Similarity 100.0%; Pred. No. 2.4e-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAGAAAGCGCTAGCATGCGCTTGTAGTGTGCTGCGAGCCCTCCAGAGACCCCTT 60
DB 59 GCAGAAAGCGCTAGCATGCGCTTGTAGTGTGCTGCGAGCCCTCCAGAGACCCCTT 118
QY 61 CCCGAGAGACCATAGTGTGCTGCGAGACCGGTGAGTACACGGGAATTGCCAGAGAC 120
DB 119 CCCGAGAGACCATAGTGTGCTGCGAGACCGGTGAGTACACGGGAATTGCCAGAGAC 178
QY 121 GGGTCTTCTTGATGATCAACCCGCTCATGCTGAGATTGGGGTGCCTCCCGGAGAC 180
DB 179 GGGTCTTCTTGATGATCAACCCGCTCATGCTGAGATTGGGGTGCCTCCCGGAGAC 238
QY 181 TCGTAGCCAGTAGTGTGCTGCGAGAAAGCGCTTGTGCTACTGCTATAGGTGCTTG 240

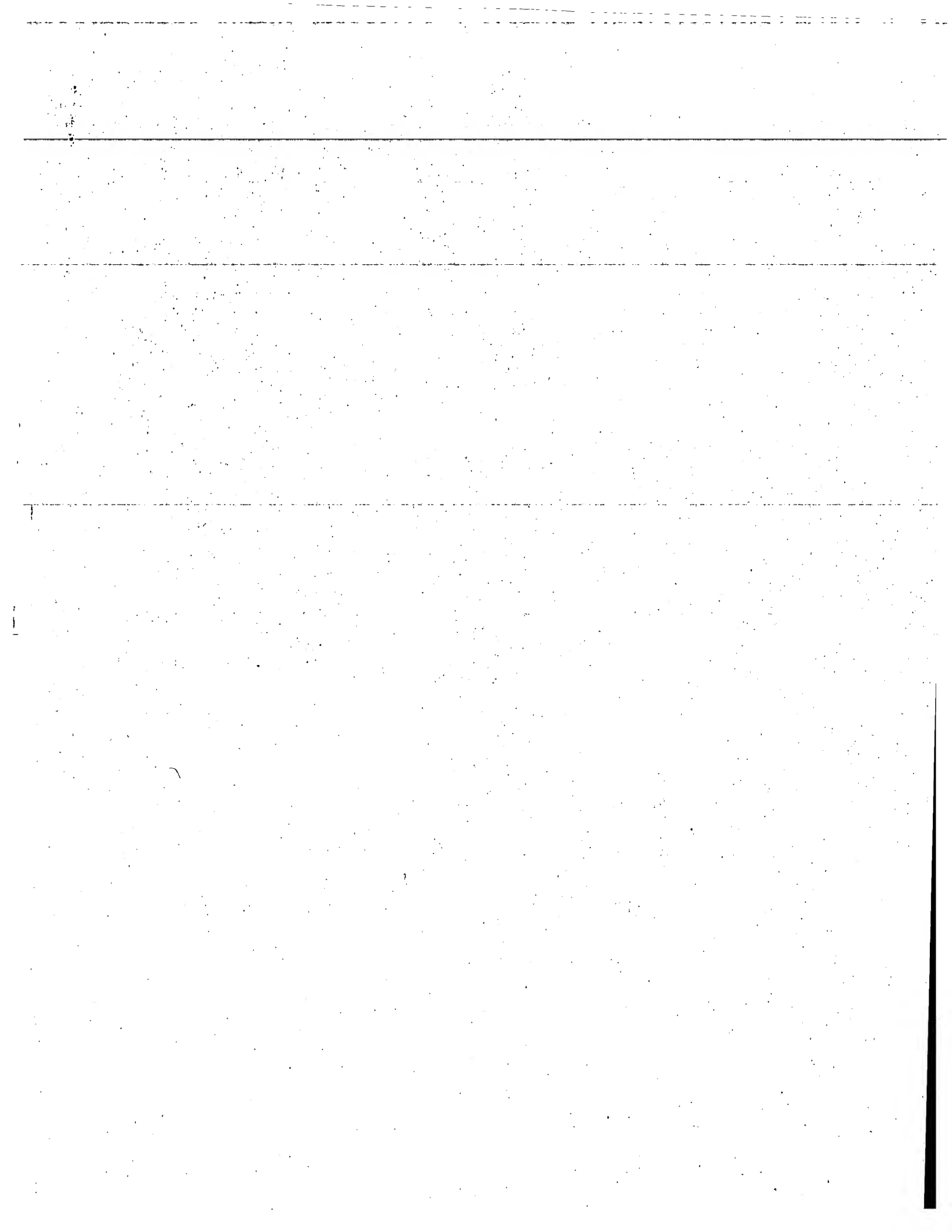
DB 239 TCGTAGCCAGTAGTGTGCTGCGAGAAAGCGCTTGTGCTACTGCTATAGGTGCTTG 298
QY 241 C 241
DB 299 C 299
RESULT 15
US-08-384-616-3
Sequence 3, Application US/08384616
Patient No. 58477101
GENERAL INFORMATION:
APPLICANT: OKAYAMA, Hiroto
APPLICANT: FUKE, Isao
APPLICANT: MORI, Chisato
APPLICANT: TAKAMIZAWA, Akahisa
APPLICANT: YOSHIDA, Itao
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
TITLE OF INVENTION: cDNA AND ANTIGEN POLYPEPTIDE
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Armstrong, Westernman, Hattori, Mclelland &
STREET: 1725 K St. N.W. Suite 1000
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384,616
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/769,996
FILING DATE: 02-OCT-1991
APPLICATION NUMBER: JP 2-167466
FILING DATE: 25-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-230921
FILING DATE: 31-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-305605
FILING DATE: 09-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/635,451
FILING DATE: 28-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Stevens-Smith, Theresa M.
REGISTRATION NUMBER: 36,281
REFERENCE/DOCKET NUMBER: 900703B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
TELEX: 440142
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1499 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA from genomic RNA
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1499
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..1499
OTHER INFORMATION: /note= "sequence = 1 - 1499 of SEQ

OTHER INFORMATION: ID: NO: 1"
US-08-384-616-3

Query Match 100.0%; Score 241; DB 2; Length 1499;
Best Local Similarity 100.0%; Pred. No. 2.4e-68;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 GCAGAAAGCGTCTAGCCATGCGGTTAGTATGATGTCGTGACGCTCCAGGACCCCCCT 60
   |||||||
Db 59 GCAGAAAGCGTCTAGCCATGCGGTTAGTATGATGTCGTGACGCTCCAGGACCCCCCT 118
   |||||||
OY 61 CCCGGAGAGCCATAGTGTCTGCGAACCGGTAGTACACCGGAAATTGCCAGGACGACC 120
   |||||||
Db 119 CCCGGAGAGCCATAGTGTCTGCGAACCGGTAGTACACCGGAAATTGCCAGGACGACC 178
   |||||||
OY 121 GGGTCCTTCTTGATCAACCCGCTCAATGCTGGAGATTGGCGCTGCCCGGAGAC 180
   |||||||
Db 179 GGGTCCTTCTTGATCAACCCGCTCAATGCTGGAGATTGGCGCTGCCCGGAGAC 238
   |||||||
OY 181 TGCTAGCCGAGTGTGTTGGTCCGAAAGGCTTGTGTAAGTGTGCTGATAGGGTCTTG 240
   |||||||
Db 239 TGCTAGCCGAGTGTGTTGGTCCGAAAGGCTTGTGTAAGTGTGCTGATAGGGTCTTG 298
   |||||||
OY 241 C 241
   |
Db 299 C 299
```

Search completed: July 10, 2003, 20:28:04
Job time : 109.773 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:13:45 ; Search time 593.333 Seconds
(without alignments)
900.761 Million cell updates/sec

Title: US-10-087-631B-3

Perfect score: 33

Sequence: 1 cgtgtactacacgttcgcagaccactatgac 33

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database :

EST:*
1: em_estb:*
2: em_esth:*
3: em_esti:*
4: em_estm:*
5: em_esto:*
6: em_estp:*
7: em_estr:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20.2	61.2	370	10	AV398524 AV398524
2	20.2	61.2	466	9	AU004535 AU004535
3	20.2	61.2	616	10	AM495169 UT-M-BH3-
4	20.2	61.2	650	10	AV398282 AV398282
5	20.2	61.2	667	10	AV398654 AV398654
6	20.2	61.2	671	10	AV398957 AV398957

7	20.2	61.2	675	10	AV398322
8	20.2	61.2	677	10	AV398498
9	20.2	61.2	677	10	AV399052
10	20.2	61.2	679	10	AV398803
11	20.2	61.2	693	10	AV398253
12	20.2	61.2	701	10	AV398450
13	20.2	61.2	713	10	AV401743
14	20.2	61.2	735	10	AV398164
15	20.2	61.2	790	10	AV398482
16	20.2	61.2	1845	11	AK005732
17	20	60.6	467	10	AM195726
18	20	60.6	772	12	BF203754
19	20	60.6	1675	12	BG025626
20	19.8	60.0	214	10	AV300075
21	19.8	60.0	421	9	AL835471
22	19.8	60.0	487	14	B0552508
23	19.8	60.0	507	9	AL834604
24	19.8	60.0	567	17	A0599676
25	19.8	60.0	589	14	C78921
26	19.8	60.0	744	10	BE260962
27	19.8	60.0	1033	11	AK018050
28	19.6	59.4	395	12	AZ244546
29	19.6	59.4	644	12	BG333148
30	19.6	59.4	792	12	BF218319
31	19.6	59.4	1134	12	BG254464
32	19.4	58.8	213	10	BB165683
33	19.4	58.8	373	9	AA953786
34	19.4	58.8	576	9	AU148437
35	19.4	58.8	581	9	AU171278
36	19.4	58.8	594	13	B0082203
37	19.4	58.8	646	13	B0111227
38	19.4	58.8	666	9	AU178284
39	19.4	58.8	668	13	BJ487077
40	19.4	58.8	670	13	BJ007208
41	19.4	58.8	675	9	AU169949
42	19.4	58.8	680	13	B0139902
43	19.4	58.8	697	13	B020967
44	19.4	58.8	734	12	BG503487
45	19.4	58.8	755	9	AU169952

ALIGNMENTS

RESULT 1
LOCUS AV398524 370 bp mRNA linear EST 05-FEB-2000
DEFINITION AV398524 Bombyx mori ovary BmPV infected; 2 hr after inoculation
Bombyx mori CDNA clone NV021911X T3, mRNA sequence.
ACCESSION AV398524
VERSION AV398524.1 GI:6902176
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.

REFERENCE
1 (bases 1 to 370)
Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.
Bombyx mori CDNA
Unpublished (2000)
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmitsa@nirs.go.jp
method: uni-directional, sequence direction; sequenced from T3 primer
(5' -> 3')

FEATURES

Project: 'Silkworm Genome Program in MAF, and Research for the
Future Program in JSPS', see 'Silkbase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
Location/Qualifiers
1..370

```

/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone_id="Bombyx mori ovary BmNPV infected; 2 hr after
inoculation"
/tissue_type="ovary"
/cell_type="BmNPV infected cell"
/dev_stage="BmNPV infected; 2 hr after inoculation"
63 a 102 c 115 g 89 t 1 others
BASE COUNT
ORIGIN

Query Match 61.2%; Score 20.2; DB 10; Length 370;
Best Local Similarity 75.8%; Pred. No. 5.3e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1 CGGTGACTCAGCGTCCGAGACCACTATGCG 33
82 CGCGGAACTCTCCGTTCCGTGCGCCCACTATGCG 114

RESULT 2
AU004535 466 bp mRNA linear EST 19-JAN-1999
LOCUS Bombyx mori p50(Daizo) Bombyx mori CDNA clone ws20577,
DEFINITION
ACCESSION AU004535.1 GI:4161906
VERSION AU004535.1
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
1 (bases 1 to 466)
Mita,K., Moriyama,M., Shimada,T., Okano,K. and Maeda,S.
Establishment of CDNA database of Bombyx mori
Unpublished (1999)
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmitsuen@irs.go.jp
PROJECT - 'CREST project by JST'
FEATURES
source
location/Qualifiers
1..466
/organism="Bombyx mori"
/strain="p50(Daizo)"
/db_xref="taxon:7091"
/clone_id="ws20577"
/clone_id="Bombyx mori p50(Daizo)"
82 a 125 c 147 g 112 t
BASE COUNT
ORIGIN

Query Match 61.2%; Score 20.2; DB 9; Length 466;
Best Local Similarity 75.8%; Pred. No. 5.7e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1 CGGTGACTCAGCGTCCGAGACCACTATGCG 33
91 CGCGGAACTCTCCGTTCCGTGCGCCCACTATGCG 123

RESULT 3
AM495169 616 bp mRNA linear EST 24-FEB-2000
LOCUS Bombyx mori p50(Daizo) Bombyx mori CDNA clone
DEFINITION
ACCESSION AM495169
VERSION AM495169.1 GI:7065415
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
1 (bases 1 to 616)
Bomaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477

JOURNAL
MEDLINE
COMMENT
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301.443.1706
Fax: 301.443.9890
Email: mst@mail.nih.gov
Oligo-dt track not found, Not I site shown in beginning of sequence
is likely internal to the message. CDNA library preparation: M.B.
Soares Lab clone distribution: Researchers may obtain BMAP CDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP CDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLY-A-NO.

FEATURES
source
location/Qualifiers

```

1..616
/organism="Mus musculus"
/strain="C57Bl/6J"
/db_xref="taxon:10090"
/clone_id="UI-M-BH3-aur-h-07-0-UI"
/clone_id="NIH_BMAP_M.S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_M.S4 library is a subtracted library of a series
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M.S4,
NIH_BMAP_M.S3.3, NIH_BMAP_M.S3.2, NIH_BMAP_M.S3.1,
NIH_BMAP_M.S2, NIH_BMAP_M.S1. The subtracted library
(NIH_BMAP_M.S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M.S3.3, NIH_BMAP_M.S3.2, and
NIH_BMAP_M.S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M.S3.3, NIH_BMAP_M.S3.2, and NIH_BMAP_M.S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M.S4 library. This procedure has been previously
described (Bomaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=NIH_BMAP_M.S4
TAG_TISSUE=basal-ganglia
TAG_SEQ=GTATAC"
BASE COUNT
ORIGIN
137 a 179 c 148 g 152 t

```

```

Query Match 61.2%; Score 20.2; DB 10; Length 616;
Best Local Similarity 75.8%; Pred. No. 6.1e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1 CGGTGACTCAGCGTCCGAGACCACTATGCG 33
165 CAGGTATCCAGAGCTCCGAGCACTATGCG 197

```

RESULT 4
AV398282 650 bp. mRNA linear. EST 05-FEB-2000
LOCUS AV398282 Bombyx mori ovary BmNPV infected; 2 hr after inoculation
DEFINITION AV398282 Bombyx mori CDNA clone NV021341 T3, mRNA sequence.
ACCESSION AV398282
VERSION AV398282.1 GI:6901934
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
1 (bases 1 to 650)
Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.
Bombyx mori CDNA
Unpublished (2000)
CONTACT: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@irs.go.jp
method: uni-directional, sequence direction: sequenced from T3 primer
(5' -> 3')

Project: 'Silkworm Genome Program in MAF, and Research for the
Future Program in JSPS', see 'Silkbase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
Location/Qualifiers
1. 650
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV021341"
/clone_lib="Bombyx mori ovary BmNPV infected; 2 hr after
inoculation"
/tissue_type="ovary"
/cell_type="BmNPV cultured cell"
/dev_stage="BmNPV infected; 2 hr after inoculation"
182 c 199 g 153 t 3 others

BASE COUNT 113 a 182 c 199 g 153 t 3 others
ORIGIN

Query Match 61.2%; Score 20.2; DB 10; Length 650;
Best Local Similarity 75.8%; Pred. No. 6.2e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 CGGTACTACCGTTCGCGACACATATGCG 33
|| ||||| ||||| ||||| ||||| ||
82 CGCGAAGCTCTCCGTTCGCGACACATATGCG 114

RESULT 5
AV399654 667 bp. mRNA linear. EST 05-FEB-2000
LOCUS AV399654 Bombyx mori ovary BmNPV infected; 12 hr after inoculation
DEFINITION AV399654 Bombyx mori CDNA clone NV120518 T3, mRNA sequence.
ACCESSION AV399654
VERSION AV399654.1 GI:6903306
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
1 (bases 1 to 667)
Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.
Bombyx mori CDNA
Unpublished (2000)
CONTACT: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@irs.go.jp
method: uni-directional, sequence direction: sequenced from T3 primer
(5' -> 3')

Project: 'Silkworm Genome Program in MAF, and Research for the
Future Program in JSPS', see 'Silkbase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
Location/Qualifiers
1. 667
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV120518"
/clone_lib="Bombyx mori ovary BmNPV infected; 12 hr after
inoculation"
/tissue_type="ovary"
/cell_type="BmNPV cultured cell"
/dev_stage="BmNPV infected; 12 hr after inoculation"
128 a 188 c 198 g 153 t 7 others

BASE COUNT 128 a 188 c 198 g 153 t 7 others
ORIGIN

Query Match 61.2%; Score 20.2; DB 10; Length 667;
Best Local Similarity 75.8%; Pred. No. 6.3e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 CGGTACTACCGTTCGCGACACATATGCG 33
|| ||||| ||||| ||||| ||||| ||
62 CGCGAAGCTCTCCGTTCGCGACACATATGCG 94

RESULT 6
AV398957 671 bp. mRNA linear. EST 05-FEB-2000
LOCUS AV398957 Bombyx mori ovary BmNPV infected; 6 hr after inoculation
DEFINITION AV398957 Bombyx mori CDNA clone NV060470 T3, mRNA sequence.
ACCESSION AV398957
VERSION AV398957.1 GI:6902609
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
1 (bases 1 to 671)
Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.
Bombyx mori CDNA
Unpublished (2000)
CONTACT: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@irs.go.jp
method: uni-directional, sequence direction: sequenced from T3 primer
(5' -> 3')

Project: 'Silkworm Genome Program in MAF, and Research for the
Future Program in JSPS', see 'Silkbase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
Location/Qualifiers
1. 671
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV060470"
/clone_lib="Bombyx mori ovary BmNPV infected; 6 hr after
inoculation"
/tissue_type="ovary"
/cell_type="BmNPV cultured cell"
/dev_stage="BmNPV infected; 6 hr after inoculation"
122 a 186 c 202 g 154 t 7 others

BASE COUNT 122 a 186 c 202 g 154 t 7 others
ORIGIN

Query Match 61.2%; Score 20.2; DB 10; Length 671;
Best Local Similarity 75.8%; Pred. No. 6.3e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 CGGTACTACCGTTCGCGACACATATGCG 33
|| ||||| ||||| ||||| ||||| ||
79 CGCGAAGCTCTCCGTTCGCGACACATATGCG 111

RESULT 7
AV398322 675 bp mRNA linear EST 05-FEB-2000
LOCUS AV398322 Bombyx mori ovary BmNPV infected; 2 hr after inoculation
DEFINITION Bombyx mori cDNA clone NV021679 T3, mRNA sequence.
ACCESSION AV398322 GI:6901974
VERSION AV398322.1
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea; Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 675)
AUTHORS Mita, K., Morimoto, M., Shimada, T., Okano, K. and Maeda, S.
TITLE Bombyx mori cDNA
JOURNAL Unpublished (2000)
COMMENT Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@nirs.go.jp
method:uni-directional, sequence direction:sequenced from T3 primer

FEATURES
source
1..675
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV021679"
/clone_lib="Bombyx mori ovary BmNPV infected; 2 hr after inoculation"
/tissue_type="ovary"
/cell_type="BmNPV infected; 2 hr after inoculation"
/dev_stage="BmNPV infected; 2 hr after inoculation"
BASE COUNT 122 a 190 c 206 g 157 t
ORIGIN

Query Match 61.2%; Score 20.2; DB 10; Length 675;
Best Local Similarity 75.8%; Pred. No. 6.3e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DB 1 CGGTGCTACTACCGTTCGCCAGACCACTATGCG 33
82 CGCGAAGCTCTCCGTCGTCGCCCACTATGCG 114

RESULT 8
AV398498 677 bp mRNA linear EST 05-FEB-2000
LOCUS AV398498 Bombyx mori ovary BmNPV infected; 2 hr after inoculation
DEFINITION Bombyx mori cDNA clone NV021881 T3, mRNA sequence.
ACCESSION AV398498
VERSION AV398498.1 GI:6902150
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea; Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 677)
AUTHORS Mita, K., Morimoto, M., Shimada, T., Okano, K. and Maeda, S.
TITLE Bombyx mori cDNA
JOURNAL Unpublished (2000)
COMMENT Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@nirs.go.jp
method:uni-directional, sequence direction:sequenced from T3 primer

(5' -> 3')
Project-"Silkworm Genome Program in MARF, and Research for the Future Program in JSPS", see 'Silkbase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
Location/Qualifiers
1..677
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV021881"
/clone_lib="Bombyx mori ovary BmNPV infected; 2 hr after inoculation"
/tissue_type="ovary"
/cell_type="BmNPV infected; 2 hr after inoculation"
/dev_stage="BmNPV infected; 2 hr after inoculation"
BASE COUNT 125 a 191 c 203 g 158 t
ORIGIN

FEATURES
source
1..677
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV060603"
/clone_lib="Bombyx mori ovary BmNPV infected; 6 hr after inoculation"
/tissue_type="ovary"
/cell_type="BmNPV infected; 6 hr after inoculation"
/dev_stage="BmNPV infected; 6 hr after inoculation"
BASE COUNT 121 a 189 c 205 g 157 t 5 others
ORIGIN

Query Match 61.2%; Score 20.2; DB 10; Length 677;
Best Local Similarity 75.8%; Pred. No. 6.3e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DB 1 CGGTGCTACTACCGTTCGCCAGACCACTATGCG 33
82 CGCGAAGCTCTCCGTCGTCGCCCACTATGCG 114

RESULT 9
AV399052 677 bp mRNA linear EST 05-FEB-2000
LOCUS AV399052 Bombyx mori ovary BmNPV infected; 6 hr after inoculation
DEFINITION Bombyx mori cDNA clone NV060603 T3, mRNA sequence.
ACCESSION AV399052
VERSION AV399052.1 GI:6902704
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea; Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 677)
AUTHORS Mita, K., Morimoto, M., Shimada, T., Okano, K. and Maeda, S.
TITLE Bombyx mori cDNA
JOURNAL Unpublished (2000)
COMMENT Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@nirs.go.jp
method:uni-directional, sequence direction:sequenced from T3 primer
(5' -> 3')
Project-"Silkworm Genome Program in MARF, and Research for the Future Program in JSPS", see 'Silkbase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
Location/Qualifiers
1..677
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV060603"
/clone_lib="Bombyx mori ovary BmNPV infected; 6 hr after inoculation"
/tissue_type="ovary"
/cell_type="BmNPV infected; 6 hr after inoculation"
/dev_stage="BmNPV infected; 6 hr after inoculation"
BASE COUNT 121 a 189 c 205 g 157 t 5 others
ORIGIN

```

RESULT 10
AV398803      679 bp  mRNA  linear  EST 05-FEB-2000
DEFINITION   AV398803 Bombyx mori ovary BmNPV infected; 6 hr after inoculation
Bombyx mori cDNA clone NV060300 T3, mRNA sequence.
ACCESSION    AV398803
VERSION      AV398803.1 GI:6902455
KEYWORDS     EST.
SOURCE       domestic silkworm.
ORGANISM     Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
1 (bases 1 to 679)
Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
Bombyx mori cDNA
Unpublished (2000)
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmitta@nirs.go.jp
method:uni-directional, sequence direction:sequenced from T3 primer
(5' -> 3')
Project-'Silkworm Genome Program in MARF, and Research for the
Future Program in JSPS', see 'Silkbase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
Location/Qualifiers
1..679
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV060300"
/clone_lib="Bombyx mori ovary BmNPV infected; 6 hr after
inoculation"
/tissue_type="ovary"
/dev_stage="BmNPV infected; 6 hr after inoculation"
BASE COUNT   121 a 189 c 205 g 158 t 6 others
ORIGIN
Query Match      61.2%; Score 20.2; DB 10; Length 679;
Best Local Similarity 75.8%; Pred. No. 6.3e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CGGTACTACACCGTTCCGACAGACCATATGCG 33
    || ||||| ||||| ||||| ||||| ||
    82 CGCGGAGCTCTCCGTCGTCGCGCCACCATATGCG 114

RESULT 11
AV398253      693 bp  mRNA  linear  EST 05-FEB-2000
DEFINITION   AV398253 Bombyx mori ovary BmNPV infected; 2 hr after inoculation
Bombyx mori cDNA clone NV021292 T3, mRNA sequence.
ACCESSION    AV398253
VERSION      AV398253.1 GI:6901905
KEYWORDS     EST.
SOURCE       domestic silkworm.
ORGANISM     Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
1 (bases 1 to 693)
Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
Bombyx mori cDNA
Unpublished (2000)
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmitta@nirs.go.jp
method:uni-directional, sequence direction:sequenced from T3 primer
(5' -> 3')
Project-'Silkworm Genome Program in MARF, and Research for the
Future Program in JSPS', see 'Silkbase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
Location/Qualifiers
1..701
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV021825"
/clone_lib="Bombyx mori ovary BmNPV infected; 2 hr after
inoculation"
/tissue_type="ovary"
/dev_stage="BmNPV infected; 2 hr after inoculation"
BASE COUNT   125 a 198 c 214 g 164 t
ORIGIN
Query Match      61.2%; Score 20.2; DB 10; Length 701;
Best Local Similarity 75.8%; Pred. No. 6.4e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CGGTACTACACCGTTCCGACAGACCATATGCG 33
    || ||||| ||||| ||||| ||||| ||

```

```

method:uni-directional, sequence direction:sequenced from T3 primer
(5' -> 3')
Project-'Silkworm Genome Program in MARF, and Research for the
Future Program in JSPS', see 'Silkbase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
Location/Qualifiers
1..693
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV021292"
/clone_lib="Bombyx mori ovary BmNPV infected; 2 hr after
inoculation"
/tissue_type="ovary"
/dev_stage="BmNPV infected; 2 hr after inoculation"
BASE COUNT   128 a 193 c 212 g 160 t
ORIGIN
Query Match      61.2%; Score 20.2; DB 10; Length 693;
Best Local Similarity 75.8%; Pred. No. 6.3e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CGGTACTACACCGTTCCGACAGACCATATGCG 33
    || ||||| ||||| ||||| ||||| ||
    89 CGCGGAGCTCTCCGTCGTCGCGCCACCATATGCG 121

RESULT 12
AV398450      701 bp  mRNA  linear  EST 05-FEB-2000
LOCUS        AV398450
DEFINITION   AV398450 Bombyx mori ovary BmNPV infected; 2 hr after inoculation
Bombyx mori cDNA clone NV021825 T3, mRNA sequence.
ACCESSION    AV398450
VERSION      AV398450.1 GI:6902102
KEYWORDS     EST.
SOURCE       domestic silkworm.
ORGANISM     Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
1 (bases 1 to 701)
Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
Bombyx mori cDNA
Unpublished (2000)
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmitta@nirs.go.jp
method:uni-directional, sequence direction:sequenced from T3 primer
(5' -> 3')
Project-'Silkworm Genome Program in MARF, and Research for the
Future Program in JSPS', see 'Silkbase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
Location/Qualifiers
1..701
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV021825"
/clone_lib="Bombyx mori ovary BmNPV infected; 2 hr after
inoculation"
/tissue_type="ovary"
/dev_stage="BmNPV infected; 2 hr after inoculation"
BASE COUNT   125 a 198 c 214 g 164 t
ORIGIN
Query Match      61.2%; Score 20.2; DB 10; Length 701;
Best Local Similarity 75.8%; Pred. No. 6.4e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CGGTACTACACCGTTCCGACAGACCATATGCG 33
    || ||||| ||||| ||||| ||||| ||

```

DB 82 CGCGAAGCTCGCTCCGTCGCGCCACTATGCG 114

RESULT 13
LOCUS AV401743
DEFINITION AV401743 Bombyx mori C108 spinning stage day-0 Bombyx mori cDNA
clone hes00104 T3, mRNA sequence.
AV401743
AV401743.1 GI:6905395
EST.
domestic silkworm.
Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
1 (bases 1 to 713)
Mita,K., Morimoto,M., Shimada,T., Okano,K. and Maeda,S.
Bombyx mori cDNA
Unpublished (2000)
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 61.2%; Score 20.2; DB 10; Length 713;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DB 57 CGCGAAGCTCGCTCCGTCGCGCCACTATGCG 89

RESULT 14
LOCUS AV398164
DEFINITION AV398164 Bombyx mori ovary BmNPV infected; 2 hr after inoculation
AV398164
AV398164.1 GI:6901816
EST.
domestic silkworm.
Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
1 (bases 1 to 735)
Mita,K., Morimoto,M., Shimada,T., Okano,K. and Maeda,S.
Bombyx mori cDNA
Unpublished (2000)
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 75.8%; Score 20.2; DB 10; Length 713;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DB 1 CGGTGTACTACCGTCCGACGACACTATGCG 33

RESULT 15
LOCUS AV398482
DEFINITION AV398482 Bombyx mori ovary BmNPV infected; 2 hr after inoculation
AV398482
AV398482.1 GI:6902134
EST.
domestic silkworm.
Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
1 (bases 1 to 790)
Mita,K., Morimoto,M., Shimada,T., Okano,K. and Maeda,S.
Bombyx mori cDNA
Unpublished (2000)
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 75.8%; Score 20.2; DB 10; Length 790;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DB 72 CGCGAAGCTCGCTCCGTCGCGCCACTATGCG 104

RESULT 15
LOCUS AV398482
DEFINITION AV398482 Bombyx mori ovary BmNPV infected; 2 hr after inoculation
AV398482
AV398482.1 GI:6902134
EST.
domestic silkworm.
Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
1 (bases 1 to 790)
Mita,K., Morimoto,M., Shimada,T., Okano,K. and Maeda,S.
Bombyx mori cDNA
Unpublished (2000)
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 61.2%; Score 20.2; DB 10; Length 735;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DB 1 CGGTGTACTACCGTCCGACGACACTATGCG 33

RESULT 15
LOCUS AV398482
DEFINITION AV398482 Bombyx mori ovary BmNPV infected; 2 hr after inoculation
AV398482
AV398482.1 GI:6902134
EST.
domestic silkworm.
Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
1 (bases 1 to 790)
Mita,K., Morimoto,M., Shimada,T., Okano,K. and Maeda,S.
Bombyx mori cDNA
Unpublished (2000)
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 61.2%; Score 20.2; DB 10; Length 790;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DB 1 CGGTGTACTACCGTCCGACGACACTATGCG 33

Email: kmita@nirs.go.jp
method: uni-directional, sequence direction: sequenced from T3 primer
(5' -> 3')
Project: 'Silkworm Genome Program in MAF, and Research for the
Future Program in JSPS', see 'SilkBase',
<http://www.db.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
Location/Qualifiers
1. 735
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="NV021198"
/clone_1lb="Bombyx mori ovary BmNPV infected; 2 hr after
inoculation"
/tissue_type="ovary"
/cell_type="BmNPV cultured cells"
/dev_stage="BmNPV infected; 2 hr after inoculation"

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 61.2%; Score 20.2; DB 10; Length 790;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DB 1 CGGTGTACTACCGTCCGACGACACTATGCG 33

RESULT 15
LOCUS AV398482
DEFINITION AV398482 Bombyx mori ovary BmNPV infected; 2 hr after inoculation
AV398482
AV398482.1 GI:6902134
EST.
domestic silkworm.
Bombyx mori
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Bombycidae; Bombyx.
1 (bases 1 to 790)
Mita,K., Morimoto,M., Shimada,T., Okano,K. and Maeda,S.
Bombyx mori cDNA
Unpublished (2000)
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 61.2%; Score 20.2; DB 10; Length 790;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DB 1 CGGTGTACTACCGTCCGACGACACTATGCG 33

Fri Jun 11 15:29:34 2003

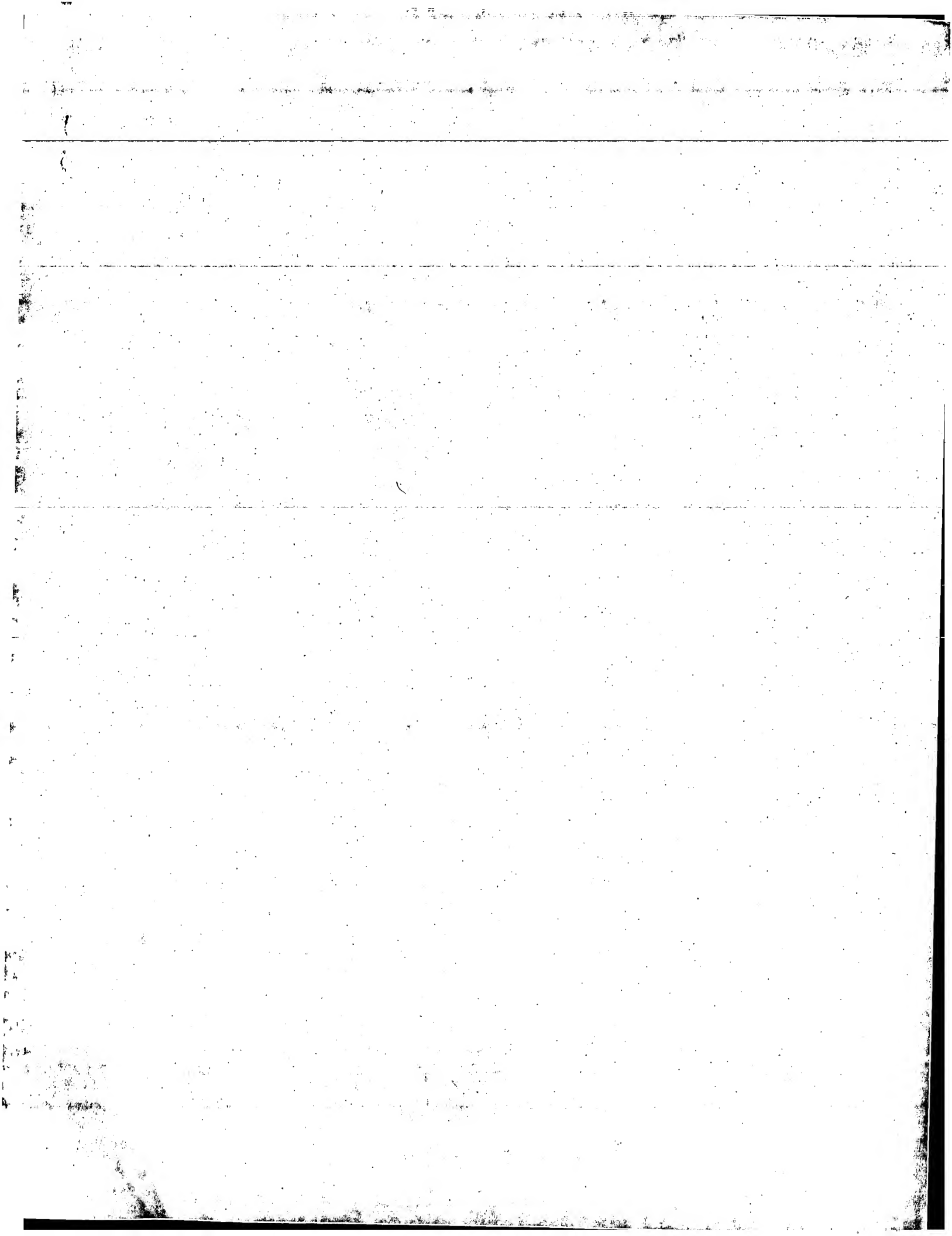
us-10-087-631b-3.rst

Page 7

Db 66 CGCGAACTCTCCGTCGTCGCGCCCACTATCGC 98

Search completed: July 11, 2003, 02:25:28
Job time : 600.333 secs

Search completed: July 11, 2003, 02:25:28
Job time : 600.333 secs



	RESULT 1
LOCUS	B1014511/c
DEFINITION	354 bp mRNA linear EST 13-JUN-2001
ACCESSION	P80-ET0208-030201-005-c04 ET0208 Homo sapiens CDNA, mRNA sequence.
VERSION	B1014511.
KEYWORDS	B1014511.1 GI:14418582
SOURCE	EST.
ORGANISM	human. Homo sapiens.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 354)
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Britones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsumura,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,K.R., Reis,L.F., de Souza,S.U. and Simpson,A.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF-expressed sequence tags.
JOURNAL MEDLINE COMMENT	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) Contact: Simpson A.J.G. Laboratory Of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel.: +55-11-2704922 Fax: +55-11-2707001

FEATURES
source

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=PM0&t2=PM0-ET0208-030201-005-c04&t3=2001-02-03&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 27
High quality sequence stop: 354
Location/Qualifiers
1. 354

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 82.8%; Score 21; DB 13; Length 354;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

ORGANISM
Human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mgc.ncl.nih.gov/
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9578 row: d column: 02
High quality sequence start: 18
High quality sequence stop: 551.
Location/Qualifiers
1. 912

FEATURES
source

Query Match
Best Local Similarity 82.8%; Score 21; DB 13; Length 354;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

ORGANISM
Human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mgc.ncl.nih.gov/
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9578 row: d column: 02
High quality sequence start: 18
High quality sequence stop: 551.
Location/Qualifiers
1. 912

FEATURES
source

Query Match
Best Local Similarity 82.8%; Score 21; DB 13; Length 354;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

ORGANISM
Human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mgc.ncl.nih.gov/
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9578 row: d column: 02
High quality sequence start: 18
High quality sequence stop: 551.
Location/Qualifiers
1. 912

FEATURES
source

Query Match
Best Local Similarity 82.8%; Score 21; DB 13; Length 354;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

ORGANISM
Human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mgc.ncl.nih.gov/
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9578 row: d column: 02
High quality sequence start: 18
High quality sequence stop: 551.
Location/Qualifiers
1. 912

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 82.8%; Score 21; DB 12; Length 912;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

ORGANISM
Human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mgc.ncl.nih.gov/
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC/DC/DTP
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12332 row: h column: 21
High quality sequence stop: 707.
Location/Qualifiers
1. 1008

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 82.8%; Score 21; DB 13; Length 1008;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

ORGANISM
Human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mgc.ncl.nih.gov/
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC/DC/DTP
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12332 row: h column: 21
High quality sequence stop: 707.
Location/Qualifiers
1. 1008

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 82.8%; Score 21; DB 13; Length 1008;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

ORGANISM
Human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mgc.ncl.nih.gov/
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC/DC/DTP
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12332 row: h column: 21
High quality sequence stop: 707.
Location/Qualifiers
1. 1008

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 82.8%; Score 21; DB 13; Length 1008;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

ORGANISM
Human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mgc.ncl.nih.gov/
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC/DC/DTP
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12332 row: h column: 21
High quality sequence stop: 707.
Location/Qualifiers
1. 1008

Eumetazoa: Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteria; Rodentia; Sclurognathi; Muridae; Murinae; Mus
1 (bases 1 to 280)
Carninec, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,

1 (bases 1 to 260)
 Akihira, S., Shibata, K., Ogawa, Y., Konno, H., Itoh, M., Aizawa, K.,
 Kariuchi, P., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Harada,
 A., Hayatsu, N., Horii, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,
 Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C.,
 Sato, K., Shibata, Y., Shigenoto, Y., Shiraki, T., Sogabe, Y., Suganarai,
 Y., Suzuki, H., Suzuki, H., Tateo, M., Tomata, Y., Tomioka, N.,
 Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,
 Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
 MIYUN MONDO, PCW.

RIKEN Mouse ESTs
Unpublished (1999)
Contact: Chle Owa

RIKEN
Genome Science Laboratory
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098

Location/Qualifiers
1. .280

Thermosensitization and thermoactivation of thermolabile enzymes by thermal shock and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

92 a 44 c 46 g 98 t

ch	68.0%;	Score 20.4;	DB 9;	Length 280;
1 Similarity	80.0%;	Pred. No. 5.1e+02;		
24; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0;

1 TGGACTCAGTCCCTTGGTCATCTCACCTTCT 30
 ||||| ||| | ||||| |||||
 8 TGGACTCAGGACTTAGGTATCTCTCCTTCT 37

AA432035 296 bp mRNA linear EST 22-MAY-1997
zw80e02.r1 Soares_testis_NHT Homo sapiens CDNA clone IMAGE:782522
''', mRNA sequence.

AA432035.1 GI:2115743
EST.
human.

1 (bases 1 to 296)
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Hom sapiens

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Giesel, G., Jost, S.,

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Thelshing, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
Washu-Merck EST Project 1997.

WashU-Merck EST Project 1997
Unpublished (1997)
Contact: Wilson RK

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63104
Tel: 314 286 1800
Fax: 314 286 1810

Location/Qualifiers
1. . 296

Email: estewartson.wisli.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information
Putative full length read
The vector to vector length is 349
Seq primer: -28ml3 rev2 ET from Amersham.

51 a	.91 c	90 g	64
------	-------	------	----

ch 68.0%; Score 20.4; DB 9; Length 296;
1 Similarity 80.0%; Pred. No. 5.2e+02;
24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

1 TGGACTCAGTCCCTTGGTGTCATCTCACCTTCT 30
||| ||| | ||||| | | ||||| |
22 TGGTCTCACCTCCCTTGGACTTGTCAACCTTGT 51

AA758968 361 bp mRNA linear EST 31-DEC-1998
 al29a05.s1 Soares_testis_NHT Homo sapiens cDNA clone 1344176 3'
 similar to contains Alu repetitive element; mRNA sequence.
 178006

EST,
human,
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 (NCBI) (bases 1 to 361)
 NC1-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo,
, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/dbp/image/image.html

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 TGGACTGACCTCTGGTCATCTCACCCTTCT 30
 ||| ||||| ||||| ||| ||||| |||
 Db 253 TGGTCTCACCCTCTGGACTGTGTCACCTTGT 282

RESULT 9
 A1601049/c 539 bp mRNA linear EST 21-APR-1999
 LOCUS 486096511.x1 486 - leaf primordia cDNA library from Hake lab Zea
 DEFINITION mays cDNA, mRNA sequence.
 ACCESSION A1601049
 VERSION A1601049.1 GI:4610210
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 539)
 REFERENCE
 AUTHORS Walbot,V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 486096 row: E column: 11.
 FEATURES
 source Location/Qualifiers
 1..539
 /organism="Zea mays"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone_lib="486 - leaf primordia cDNA library from Hake
 lab"
 /tissue_type="leaf primordia"
 /dev_stage="P7-P11 leaf"
 /lab_host="E.coli XL1-Blue MFR"
 /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
 library."
 BASE COUNT 166 a 92 c 146 g 134 t 1 others

Query Match 68.0%; Score 20.4; DB 9; Length 539;
 Best Local Similarity 80.0%; Pred. No. 6.6e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 TGGACTGACCTCTGGTCATCTCACCCTTCT 30
 ||||| ||| ||| ||||| ||||| |||||
 Db 340 TGGACACGGTACATGTCATCTCCTCTCT 311

RESULT 10
 B0328104 542 bp mRNA linear EST 17-MAY-2002
 LOCUS MRL-RT0079-151200-003-B02 RT0079 Homo sapiens cDNA, mRNA sequence.
 DEFINITION B0328104
 ACCESSION B0328104.1 GI:20946200
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 542)
 REFERENCE
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Britones,M.R.,
 Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

TITLE
 JOURNAL
 MEDLINE
 COMMENT

FEATURES
 source Location/Qualifiers
 1..542
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="RT0079"
 /dev_stage="Adult"
 /note="Organ: kidney,tumor; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 90 a 194 c 136 g 119 t 3 others

Query Match 68.0%; Score 20.4; DB 14; Length 542;
 Best Local Similarity 80.0%; Pred. No. 6.6e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 TGGACTGACCTCTGGTCATCTCACCCTTCT 30
 ||| ||||| ||||| ||| ||||| |||
 Db 479 TGGTCTCACCCTCTGGACTGTGTCACCTTGT 508

RESULT 11
 BG383655 582 bp mRNA linear EST 12-MAR-2001
 LOCUS 301851 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
 DEFINITION BG383655
 ACCESSION BG383655.1 GI:13308127
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 582)
 REFERENCE
 AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
 Stone,R.T., Heaton,M.P., Grose,W.M., Bennett,G.A., Laegreid,W.W.
 and Keefe,J.W.
 TITLE Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine
 JOURNAL Unpublished (2000)
 COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt-trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -mismatch 12 options.

Fri Jul 11 15:29:35 2003

Search completed: July 11, 2003, 02:25:36
Job time : 547.393 secs

us-10-087-631b-4.rst

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:28 : Search time 14.7573 seconds
(without alignments)
685.785 Million cell updates/sec

Title: US-10-087-631b-3

Perfect score: 33

Sequence: 1 cgggtactaccgccttcgcagacactatggc 33

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_MA:*
1: /cgn2_6/p/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/p/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/p/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/p/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/p/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/p/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	66.7	120	2	US-08-690-495-26
2	22	66.7	120	2	US-08-690-494-26
3	22	66.7	120	2	PCT-US95-07671-26
4	22	66.7	145	2	US-08-690-495-18
5	22	66.7	145	2	US-08-690-494-18
6	22	66.7	145	5	PCT-US95-07671-18
7	22	66.7	177	2	US-08-256-5688-55
8	22	66.7	177	2	US-08-256-5688-56
9	22	66.7	177	2	US-08-256-5688-57
10	22	66.7	177	2	US-08-256-5688-58
11	22	66.7	177	2	US-08-256-5688-61
12	22	66.7	177	2	US-08-256-5688-62
13	22	66.7	177	2	US-08-256-5688-63
14	22	66.7	177	2	US-08-256-5688-64
15	22	66.7	177	2	US-08-256-5688-65
16	22	66.7	177	2	US-08-256-5688-66
17	22	66.7	177	2	US-08-256-5688-67
18	22	66.7	177	2	US-08-256-5688-68
19	22	66.7	177	2	US-08-256-5688-69
20	22	66.7	177	2	US-08-256-5688-70
21	22	66.7	177	2	US-08-256-5688-72
22	22	66.7	177	2	US-08-256-5688-73
23	22	66.7	177	2	US-08-256-5688-74
24	22	66.7	177	2	US-08-256-5688-75
25	22	66.7	177	2	US-08-256-5688-77
26	22	66.7	177	2	US-08-256-5688-78
27	22	66.7	177	2	US-08-256-5688-79

c 28	22	66.7	177	2	US-08-256-5688-80	Sequence 80, Appl
c 29	22	66.7	177	4	US-09-038-3698-55	Sequence 55, Appl
c 30	22	66.7	177	4	US-09-038-3698-56	Sequence 56, Appl
c 31	22	66.7	177	4	US-09-038-3698-57	Sequence 57, Appl
c 32	22	66.7	177	4	US-09-038-3698-58	Sequence 58, Appl
c 33	22	66.7	177	4	US-09-038-3698-61	Sequence 61, Appl
c 34	22	66.7	177	4	US-09-038-3698-62	Sequence 62, Appl
c 35	22	66.7	177	4	US-09-038-3698-63	Sequence 63, Appl
c 36	22	66.7	177	4	US-09-038-3698-64	Sequence 64, Appl
c 37	22	66.7	177	4	US-09-038-3698-65	Sequence 65, Appl
c 38	22	66.7	177	4	US-09-038-3698-66	Sequence 66, Appl
c 39	22	66.7	177	4	US-09-038-3698-67	Sequence 67, Appl
c 40	22	66.7	177	4	US-09-038-3698-68	Sequence 68, Appl
c 41	22	66.7	177	4	US-09-038-3698-69	Sequence 69, Appl
c 42	22	66.7	177	4	US-09-038-3698-70	Sequence 70, Appl
c 43	22	66.7	177	4	US-09-038-3698-72	Sequence 72, Appl
c 44	22	66.7	177	4	US-09-038-3698-73	Sequence 73, Appl
c 45	22	66.7	177	4	US-09-038-3698-74	Sequence 74, Appl

ALIGNMENTS

RESULT 1
US-08-690-495-26
: Sequence 26, Application US/08690495
: Patent No. 5876924
: GENERAL INFORMATION:
: APPLICANT: Zhang, David Y., Brandwein, Margaret
: TITLE OF INVENTION: NUCLEIC ACID AMPLIFICATION METHOD:
: TITLE OF INVENTION: HYBRIDIZATION SIGNAL AMPLIFICATION METHOD (HSAM)
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
: STREET: 30 Rockefeller Plaza
: CITY: New York
: STATE: NY
: COUNTRY: USA
: ZIP: 10112-0228
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Fasteo Version #1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/690,495
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: MacLeod, Janet M.
: REGISTRATION NUMBER: 35,263
: REFERENCE/DOCKET NUMBER: 29545-A-PCT/USA-A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-765-2519
: TELEFAX: 212-765-2519
: INFORMATION FOR SEQ ID NO: 26:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 120 base pairs
: TYPE: nucleic acid
: TOPOLOGY: linear
: STRANDEDNESS: single
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 1..120
: US-08-690-495-26
Query Match 66.7%; Score 22; DB 2; Length 120;
Best local similarity 97.1%; Pred. No. 0.27;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
1 CGGTGACTCACCGTTCGCGAGACACTATGGC 33
|||||

Db 64 CGGTGTAACACCGGTTCCGACAGACCACTATGGC 97

RESULT 2

Sequence 26, Application US/08690494
Patent No. 5942391

GENERAL INFORMATION:

APPLICANT: Zhang, David Y., Brandwein, Margaret
TITLE OF INVENTION: NUCLEIC ACID AMPLIFICATION METHOD:
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESS: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
ZIP: 10112-0228COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy diskCOMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Fasteq Version #1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/690,494

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MacLeod, Janet M.

REGISTRATION NUMBER: 35,263

REFERENCE/DOCKET NUMBER: 29545-A-PCT/USA-B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-408-2597

TELEFAX: 212-765-2519

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 120 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: misc.feature

LOCATION: 1..120

US-08-690-494-26

Query Match

Best Local Similarity 66.7%; Score 22; DB 2; Length 120;

Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 64 CGGTGTAACACCGGTTCCGACAGACCACTATGGC 97

PCT-US95-07671-26

Sequence 26, Application PC/TUS9507671

GENERAL INFORMATION:

APPLICANT: Zhang, David Y.
TITLE OF INVENTION: LIGATION-DEPENDENT AMPLIFICATION FOR THE
NUMBER OF SEQUENCES: 31
DETECTION OF INFECTIOUS PATHOGENS AND ABNORMAL GENES
CORRESPONDENCE ADDRESS:
ADDRESS: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
ZIP: 10112-0228

COUNTRY: USA

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/07671

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Seide, Rochelle K.

REGISTRATION NUMBER: 32,300

REFERENCE/DOCKET NUMBER: 29545-A-PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-408-2626

TELEFAX: 212-765-2519

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 120 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: misc.feature

LOCATION: 1..120

PCT-US95-07671-26

Query Match

Best Local Similarity 66.7%; Score 22; DB 5; Length 120;

Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 64 CGGTGTAACACCGGTTCCGACAGACCACTATGGC 97

RESULT 4

US-08-690-495-18

Sequence 18, Application US/08690495

Patent No. 5876924

GENERAL INFORMATION:

APPLICANT: Zhang, David Y., Brandwein, Margaret
TITLE OF INVENTION: NUCLEIC ACID AMPLIFICATION METHOD:
NUMBER OF SEQUENCES: 42
HYBRIDIZATION SIGNAL AMPLIFICATION METHOD (HSAM)
CORRESPONDENCE ADDRESS:
ADDRESS: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
ZIP: 10112-0228

COUNTRY: USA

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Fasteq Version #1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/690,495

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MacLeod, Janet M.

REGISTRATION NUMBER: 35,263

REFERENCE/DOCKET NUMBER: 29545-A-PCT/USA-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-408-2597

TELEFAX: 212-765-2519

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 145 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..145
US-08-690-495-18

Query Match 66.7%; Score 22; DB 2; Length 145;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 33
|||||
DB 72 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 105

RESULT 5
US-08-690-494-18
Sequence 18, Application US/08690494
Patent No. 5942391

GENERAL INFORMATION:

APPLICANT: Zhang, David Y., Brandwein, Margaret
TITLE OF INVENTION: NUCLEIC ACID AMPLIFICATION METHOD:
TITLE OF INVENTION: HYBRIDIZATION SIGNAL AMPLIFICATION METHOD (HSAM)
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10112-0228

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Fasteq Version #1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/690,494
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: MacLeod, Janet M.
REGISTRATION NUMBER: 35,263
REFERENCE/DOCKET NUMBER: 29545-A-PCT/USA-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2597
TELEFAX: 212-765-2519

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:
LENGTH: 145 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..145
US-08-690-494-18

Query Match 66.7%; Score 22; DB 2; Length 145;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 33
|||||
DB 72 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 105

RESULT 6
PCT-US95-07671-18
Sequence 18, Application PC/TUS9507671
GENERAL INFORMATION:
APPLICANT: Zhang, David Y.

TITLE OF INVENTION: LIGATION-DEPENDENT AMPLIFICATION FOR THE
DETECTION OF INFECTIOUS PATHOGENS AND ABNORMAL GENES
TITLE OF INVENTION: DETECTION OF INFECTIOUS PATHOGENS AND ABNORMAL GENES
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10112-0228

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07671
FILING DATE:
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Seide, Rochelle K.
REGISTRATION NUMBER: 32,300
REFERENCE/DOCKET NUMBER: 29545-A-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2626
TELEFAX: 212-765-2519
INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 145 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..145
PCT-US95-07671-18

Query Match 66.7%; Score 22; DB 5; Length 145;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 33
|||||
DB 72 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 105

RESULT 7

US-08-256-568B-55/C
Sequence 55, Application US/08256568B
Patent No. 5846704

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;
APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO
TITLE OF INVENTION: PROCESS FOR TYPING OF HCV
TITLE OF INVENTION: ISOLATES
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCIT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410.004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IMMEDIATE SOURCE:
CLONE: jp62
POSITION IN GENOME:
MAP POSITION: 5' untranslated region
US-08-256-568B-55

Query Match 66.7%; Score 22; DB 2; Length 177;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

DB 1 CGGTGTACTCACC-GTTCGCGAGACCATATGCG 33
73 CGGTGTACTCACC-GTTCGCGAGACCATATGCG 40

RESULT 8
US-08-256-568B-56/c
Sequence 56, Application US/08256568B
Patent No. 5846704
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;
APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO
TITLE OF INVENTION: PROCESS FOR TYPING OF HCV
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992

ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410.004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IMMEDIATE SOURCE:
CLONE: gp81
POSITION IN GENOME:
MAP POSITION: 5' untranslated region
US-08-256-568B-56

Query Match 66.7%; Score 22; DB 2; Length 177;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

DB 1 CGGTGTACTCACC-GTTCGCGAGACCATATGCG 33
73 CGGTGTACTCACC-GTTCGCGAGACCATATGCG 40

RESULT 9
US-08-256-568B-57/c
Sequence 57, Application US/08256568B
Patent No. 5846704
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;
APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO
TITLE OF INVENTION: PROCESS FOR TYPING OF HCV
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410.004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:

LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: bt96
POSITION IN GENOME:
MAP POSITION: 5' untranslated region
US-08-256-568B-57

Query Match 66.7%; Score 22; DB 2; Length 177;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 33
DB 73 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 40

RESULT 10
US-08-256-568B-58/c
Sequence 58, Application US/08256568B
Patent No. 5846704
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;
APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO
TITLE OF INVENTION: PROCESS FOR TYPING OF HCV
TITLE OF INVENTION: ISOLATES
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410,004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: bu79
POSITION IN GENOME:
MAP POSITION: 5' untranslated region

US-08-256-568B-58

Query Match 66.7%; Score 22; DB 2; Length 177;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 33
DB 73 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 40

RESULT 11
US-08-256-568B-61/c
Sequence 61, Application US/08256568B
Patent No. 5846704
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;
APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO
TITLE OF INVENTION: PROCESS FOR TYPING OF HCV
TITLE OF INVENTION: ISOLATES
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410,004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: be82 (also referred to as be99)
POSITION IN GENOME:
MAP POSITION: 5' untranslated region
US-08-256-568B-61

Query Match 66.7%; Score 22; DB 2; Length 177;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 33
DB 73 CGGTGTACTCACC-GTTCGCGAGACCACTATGGC 40

RESULT 12

US-08-256-568B-62/c
Sequence 62, Application US/08256568B
Patent No. 5846704

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;
APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO
TITLE OF INVENTION: PROCESS FOR TYPING OF HC
TITLE OF INVENTION: ISOLATES
NUMBER OF SEQUENCES: 97

CORRESPONDENCE ADDRESS:

ADDRESS: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994

CLASSIFICATION:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993

APPLICATION NUMBER: EP/93/402,129.6

FILING DATE: 31-AUG-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992

ATTORNEY/AGENT INFORMATION:

NAME: CHARLES A. MUSERLIAN

REGISTRATION NUMBER: 19,683

REFERENCE/DOCKET NUMBER: 410.004

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 661-8000

TELEFAX: (212) 661-8002

INFORMATION FOR SEQ ID NO: 62:

SEQUENCE CHARACTERISTICS:

LENGTH: 177 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

IMMEDIATE SOURCE:

CLONE: be90

POSITION IN GENOME:

MAP POSITION: 5' untranslated region

US-08-256-568B-62

Query Match

Best Local Similarity 66.7%; Score 22; DB 2; Length 177;

Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

DB 73 CGGTGACTACCGGTTCCGACGACCACTATGCG 40

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

TITLE OF INVENTION: PROCESS FOR TYPING OF HC
TITLE OF INVENTION: ISOLATES
NUMBER OF SEQUENCES: 97

CORRESPONDENCE ADDRESS:

ADDRESS: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994

CLASSIFICATION:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993

APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992

ATTORNEY/AGENT INFORMATION:

NAME: CHARLES A. MUSERLIAN

REGISTRATION NUMBER: 19,683

REFERENCE/DOCKET NUMBER: 410.004

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 661-8000

TELEFAX: (212) 661-8002

INFORMATION FOR SEQ ID NO: 63:

SEQUENCE CHARACTERISTICS:

LENGTH: 177 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

IMMEDIATE SOURCE:

CLONE: be91

POSITION IN GENOME:

MAP POSITION: 5' untranslated region

US-08-256-568B-63

Query Match

Best Local Similarity 66.7%; Score 22; DB 2; Length 177;

Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

DB 73 CGGTGACTACCGGTTCCGACGACCACTATGCG 40

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

CGGTGACTACCGGTTCCGACGACCACTATGCG 33

ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410.004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IMMEDIATE SOURCE:
CLONE: be92
POSITION IN GENOME:
MAP POSITION: 5' untranslated region
US-08-256-568B-64

Query Match 66.7%; Score 22; DB 2; Length 177;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGACTCACC-GTTCGCGACGACCATATGCG 33
DB 73 CGGTGACTCACC-GTTCGCGACGACCATATGCG 40

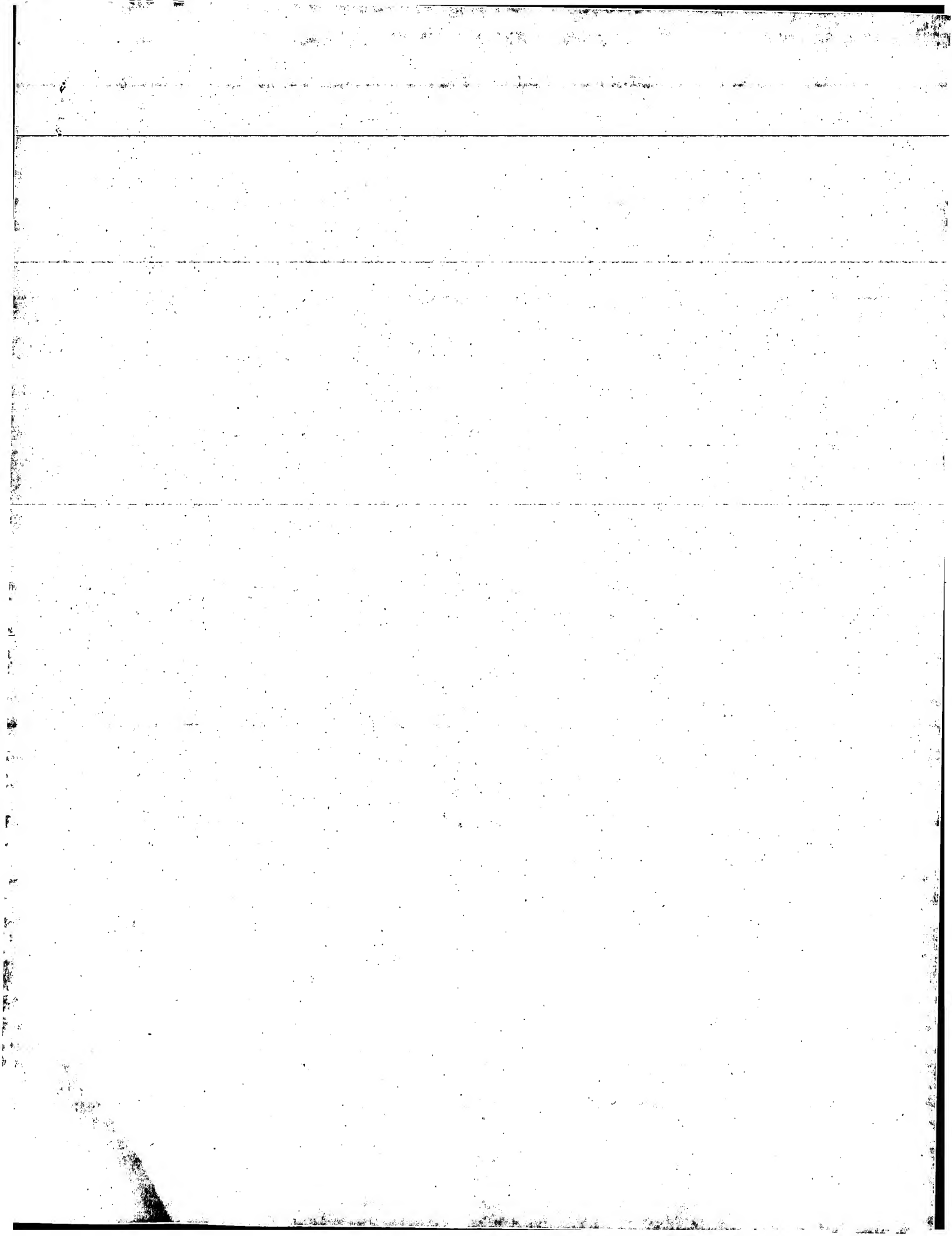
RESULT 15
US-08-256-568B-65/C
Sequence 65, Application US/08256568B
Patent No. 5846704
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT; STUYVER, LIEVEN;
APPLICANT: ROSSAU, RUDI; VAN HEUVERSWYN, HUGO
TITLE OF INVENTION: PROCESS FOR TYPING OF HCV
TITLE OF INVENTION: ISOLATES
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,568B
FILING DATE: 18-JUL-1994

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/03325
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/93/402,129.6
FILING DATE: 31-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP/92/403,222.0
FILING DATE: 27-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410.004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IMMEDIATE SOURCE:
CLONE: be93
POSITION IN GENOME:
MAP POSITION: 5' untranslated region
US-08-256-568B-65

Query Match 66.7%; Score 22; DB 2; Length 177;
Best Local Similarity 97.1%; Pred. No. 0.28;
Matches 33; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGGTGACTCACC-GTTCGCGACGACCATATGCG 33
DB 73 CGGTGACTCACC-GTTCGCGACGACCATATGCG 40

Search completed: July 10, 2003, 20:27:12
Job time: 15.7573 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:13:45 ; Search time 433.13 Seconds
(without alignments)
900.761 Million cell updates/sec

Title: US-10-087-631b-10.
Sequence: 1 gcagaaagcgtctacgcatg.....ctgcctgatatggtgcttc 241

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estb:*
2: em_esthum:*
3: em_estnu:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
C 1	35	14.5	534 9	AL798317 AL798317
C 2	35	14.5	557 9	AL647223 AL647223
C 3	35	14.5	572 9	AL636391 AL636391
C 4	35	14.5	577 9	AL784104 AL784104
C 5	34.4	14.3	928 10	BE249938 BE249938
C 6	34.2	14.2	952 9	AL563471 AL563471

C 7	34	14.1	422 9	AA172136
C 8	33.4	13.9	706 10	BB628805
C 9	33.2	13.8	774 13	BI153642
C 10	32.8	13.6	468 14	BM738887
C 11	32.8	13.6	575 12	BE888083
C 12	32.8	13.6	655 13	BI765955
C 13	32.8	13.6	925 17	CNS0091P
C 14	32.8	13.6	1331 12	BF306437
C 15	32.8	13.6	2670 11	BC014080
C 16	32.6	13.5	713 17	AG008830
C 17	32.4	13.4	349 17	AA431550
C 18	32	13.3	438 9	AA069278
C 19	32	13.3	637 13	BJ484211
C 20	32	13.3	667 13	BJ485696
C 21	32	13.3	669 10	AV945501
C 22	32	13.3	669 13	BJ483757
C 23	32	13.3	671 13	BJ483019
C 24	32	13.3	681 13	BJ482842
C 25	32	13.3	897 12	BG031227
C 26	32	13.3	984 12	BG420578
C 27	31.8	13.2	983 12	BF684127
C 28	31.6	13.1	444 9	AA633661
C 29	31.6	13.1	705 17	AG008799
C 30	31.6	13.1	922 13	BI151067
C 31	31.6	13.1	925 17	CNS0091P
C 32	31.6	13.1	1048 14	BM810632
C 33	31.4	13.0	997 13	BM459695
C 34	31.4	13.0	1047 10	BE616401
C 35	31.4	13.0	1138 14	BM921090
C 36	31.2	12.9	371 12	BF508846
C 37	31.2	12.9	549 10	BE424949
C 38	31.2	12.9	593 17	TA164C1Q
C 39	31.2	12.9	600 17	AO538955
C 40	31	12.9	177 17	AZ113390
C 41	31	12.9	465 10	BB854666
C 42	31	12.9	474 10	AM473754
C 43	31	12.9	651 9	AL043622
C 44	31	12.9	734 17	AO160900
C 45	31	12.9	838 13	BI689033

ALIGNMENTS

RESULT 1
AL798317/c 534 bp mRNA linear EST 27-JUN-2002
LOCUS AL798317 XGC-neurula silurana tropicalis cDNA clone TNeu13m07 5',
DEFINITION mRNA sequence.
ACCESSION AL798317
VERSION AL798317.1 GI:21584021
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
Xenopodinae; Silurana.
REFERENCE 1 (bases 1 to 534)
AUTHORS Taylor,R., Ashurst,J.L., Croning,M.D.R., Zorn,A.M. and Rogers,J.
TITLE Sanger Xenopus tropicalis EST project 2002
JOURNAL Unpublished (2001)
COMMENT Contact: Taylor R
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TNeu13m07.plcsp6
Sequencing primer: Plcsp6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1..534
/organism="Silurana tropicalis"

FEATURES

source

/db.xref="taxon:8364"
/clone="Tneu13m07"
/dev_stage="neurula"
/lab_host="Escherichia coli DH10B"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."
104 a 171 c 175 g 84 t

BASE COUNT
Origin
Query Match
Best Local Similarity 14.5%; Score 35; DB 9; Length 534;
Matches 86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 17 CATGCGTTAGTATGATGTCGTCGACGCTCCAGAGACCCCTCCCGGAGAGCCATAG 76
DB 313 CATGGCGGTACTACAGGCACAGGCGATCTCAGACAGGCGACGAGACTGGCGCGACGGGA 254
QY 77 TGTCTGCGGAACCGGTGATGATACACCGGAATTGCCAGACGACCGGCTCTTCTTGAT 136
DB 253 TGTCTGCGGAAGAGATGCTGACGCTGATAGAGCGGGCTGCGGGGATTTTGCACGGCT 194
QY 137 CAACCCGCTCAATGCTGAGATTTGGGCGTCCCGCCGAGACTGCTAGC 187
DB 193 CTAGCCCTTACACTCCCGGGGATTCAGCTACGACCGGCGCTTGTGCG 143

RESULT 2
AL647223/c
LOCUS AL647223 XGC-gastrula Silurana tropicalis cDNA clone Tgas026a22 5'
DEFINITION mRNA sequence.
ACCESSION AL647223
VERSION AL647223.1 GI:17655336
KEYWORDS EST.
SOURCE Western clawed frog
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Silurana.
1. (bases 1 to 577)
Huckle, E., Taylor, R., Ashurst, J. L., Zorn, A. M. and Rogers, J.
Sanger Xenopus tropicalis EST project 2001 (10_2001)
Unpublished (2001)
Contact: Huckle E
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE.ID: Tgas026a22.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1. 557

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source
/organism="Silurana tropicalis"
/db.xref="taxon:8364"
/clone="Tgas026a22"
/clone_1id="XGC-gastrula"
/dev_stage="gastrula (stages 10.5-13 mixed)"
/lab_host="Escherichia coli XL1-blue"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."
105 a 179 c 186 g 87 t

BASE COUNT
Origin
Query Match
Best Local Similarity 14.5%; Score 35; DB 9; Length 557;
Matches 86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Matches 86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 17 CATGCGTTAGTATGATGTCGTCGACGCTCCAGAGACCCCTCCCGGAGAGCCATAG 76
DB 313 CATGGCGGTACTACAGGCACAGGCGATCTCAGACAGGCGACGAGACTGGCGCGACGGGA 254
QY 77 TGTCTGCGGAACCGGTGATGATACACCGGAATTGCCAGACGACCGGCTCTTCTTGAT 136
DB 253 TGTCTGCGGAAGAGATGCTGACGCTGATAGAGCGGGCTGCGGGGATTTTGCACGGCT 194
QY 137 CAACCCGCTCAATGCTGAGATTTGGGCGTCCCGCCGAGACTGCTAGC 187
DB 193 CTAGCCCTTACACTCCCGGGGATTCAGCTACGACCGGCGCTTGTGCG 143

RESULT 3
AL636391/c
LOCUS AL636391 XGC-neurula Silurana tropicalis cDNA clone Tneu021f04 5'
DEFINITION mRNA sequence.
ACCESSION AL636391
VERSION AL636391.1 GI:16788370
KEYWORDS EST.
SOURCE Western clawed frog
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Silurana.
1. (bases 1 to 572)
Huckle, E., Taylor, R., Ashurst, J. L., Zorn, A. M. and Rogers, J.
Sanger Xenopus tropicalis EST project 2001 (10_2001)
Unpublished (2001)
Contact: Huckle E
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE.ID: Tneu021f04.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1. 572

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source
/organism="Silurana tropicalis"
/db.xref="taxon:8364"
/clone="Tneu021f04"
/clone_1id="XGC-neurula"
/dev_stage="neurula"
/lab_host="Escherichia coli DH10B"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."
98 a 180 c 199 g 95 t

BASE COUNT
Origin
Query Match
Best Local Similarity 14.5%; Score 35; DB 9; Length 572;
Matches 86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 17 CATGCGTTAGTATGATGTCGTCGACGCTCCAGAGACCCCTCCCGGAGAGCCATAG 76
DB 268 CATGGCGGTACTACAGGCACAGGCGATCTCAGACAGGCGACGAGACTGGCGCGACGGGA 209
QY 77 TGTCTGCGGAACCGGTGATGATACACCGGAATTGCCAGACGACCGGCTCTTCTTGAT 136
DB 208 TGTCTGCGGAAGAGATGCTGACGCTGATAGAGCGGGCTGCGGGGATTTTGCACGGCT 149
QY 137 CAACCCGCTCAATGCTGAGATTTGGGCGTCCCGCCGAGACTGCTAGC 187
DB 148 CTAGCCCTTACACTCCCGGGGATTCAGCTACGACCGGCGCTTGTGCG 98

LOCUS	AL784104/c	577 bp	mrna	linear	EST 25-JUN-2002
DEFINITION	AL784104 XGC-gastrula Silurana tropicalis cDNA clone Tgas073h08 5', mRNA sequence.				
ACCESSION	AL784104				
VERSION	AL784104.1 GI:21569808				
KEYWORDS	EST.				
SOURCE	western clawed frog.				
ORGANISM	Silurana tropicalis				
REFERENCE	Taylor, R., Ashurst, J.L., Croning, M.D.R., Zorn, A.M. and Rogers, J.				
AUTHORS	Sanger Xenopus tropicalis EST project 2002				
TITLE	Unpublished (2001)				
JOURNAL	Contact: Taylor R				
COMMENT	Sanger Centre Hinxton, Cambridgeshire, CB10 1SA, UK Email: troyesanger.ac.uk Sanger Xenopus tropicalis EST project 2001 TROPICALIS_SEQUENCE_ID: Tgas073h08.plksp6 Sequencing primer: PlkSP6 This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn. Location/Qualifiers 1..577 /organism="Silurana tropicalis" /db_xref="taxon:8364" /clone="Tgas073h08" /clone_lib="XGC-gastrula" /dev_stage="gastrula (stages 10.5-13 mixed)" /lab_host="Escherichia coli XL1-blue" /note="Vector: pCS107; Site.1: EcoRI; Site.2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."				
FEATURES	Source				
BASE COUNT	108 a 184 c 194 g 91 t				
ORIGIN					
Query Match	14.5% Score 35; DB 9; Length 577;				
Best Local Similarity	50.3% Pred. No. 20;				
Matches	86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;				
QY	17 CATGCGCTTACTATGAGTGTGTCGTCAGCGCTCCAGACCCCTCCCGGAGAGCCATAG 76				
Db	321 CATGCGGCTACTACACGACACAGGCGCATCCACAGAGGACGGAACCTGGCGACGCGGA 262				
OY	77 TGTGTCGGGAGACCGGTGAGTAGTACACCGGAATTCGACAGACGACGCGGTCTTTTGGAT 136				
Db	261 TGTCTCCGAGAGATGATCGTGAACCTCGATAGACGGGGCTCGCGGATTTTGACAGGCT 202				
OY	137 CAACCCGCTCATGCTCGAGAGATTTGGGCGGCCCGCCGCGAGACTGCTAGC 187				
Db	201 CTAGCCCTTACATCCCGGGGATTCAGACCTTACGACGCGGGCTTGTGGC 151				
RESULT 5					
LOCUS	BE249938/c				
DEFINITION	BE249938 928 bp mrna linear EST 13-JUL-2000				
ACCESSION	600942989F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:2959391 5', mRNA sequence.				
VERSION	BE249938				
KEYWORDS	BE249938.1 GI:9120041				
SOURCE	EST.				
ORGANISM	human.				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	NIH-MGC http://mgc.ncl.nih.gov/.				

TITLE	JOURNAL
National Institutes of Health, Mammalian Gene Collection (MGC)	
Unpublished (1999)	
Contact: Robert Strausberg, Ph.D.	
Email: cgabbs-remail.nih.gov	
Tissue Procurement: ATCC	
CNA Library Preparation: Ling Hong/Rubin Laboratory	
DNA Sequencing by: Incyte Genomics, Inc.	
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov	
Plate: L1CM51 row: a column: 24	
High quality sequence start: 2	
High quality sequence stop: 758.	
Location/Qualifiers	
1. 928	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:2959391"	
/clone_11b="NH. MGC.15"	
/tissue_type="adenocarcinoma cell line"	
/lab_host="DH10B (phage-resistant)"	
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-AT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"	
BASE COUNT	
223 a 230 c 319 g 156 t	
ORIGIN	
Query Match	14.3%; Score 34.4; DB 10; Length 928;
Best Local Similarity	48.5%; Pred. No. 33;
Matches	95; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
OY	40 GCAGCCTCCAGAGACCCCCCTCCCGGGAGAGACCCTAGTGTCTCGGAGACCGGTGAGAC 99
DB	
OY	819 GCACCTCCACGCGCTCCATCCCAATGCGCCCGAGAGACCTCGGCTTGAGACCTCAA 760
OY	100 ACCGGAATTCGCAGAGACGCGGGGCTCTTCTTGATCAACCCGCTCAATGCTGGAGAT 159
DB	759 TCCAGCCGAGACCGGATGTCGGGTTCATCTAGAGATTCACATCTTTGGCCGCCAGGTAT 700
OY	160 TTGGGCGGTGCCCCCGCGAGACTGCTAGCCGAGTAGTGTGGGTGCGGAAAGGCTTGTGG 219
DB	699 CCCCCTGTTCCACGACGAGGCTCGACAGTCTCTACTATTGATGCTACATGCTCTGAC 640
OY	220 TACTGCTGATAGGCT 235
DB	639 CTCAGCCGCTGCGGT 624
RESULT 6	
LOCUS	AL563471 952 bp mRNA linear EST 16-FEB-2001
DEFINITION	AL563471 LFI.NFL001.NBC4 Homo sapiens cDNA clone CSDD0006YF10 3
ACCESSION	AL563471
VERSION	AL563471.1 GI:12912895
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: segrete@genoscope.cns.fr , Web : www.genoscope.cns.fr .

FEATURES

Location/Qualifiers
1. 952

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CSODD006710"
/clone_lib="VRI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"

/note="Organ: brain; Vector: PCMVSPORT 6, 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact: Feng Liang life technologies, a division of invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com"

BASE COUNT

224 a 283 c 247 g 184 t 14 others

Query Match 14.1%; Score 34.2; DB 9; Length 952;
Best Local Similarity 60.7%; Pred. No. 38;
Matches 54; Conservative 1; Mismatches 34; Indels 0; Gaps 0;

OY

47 CCAGGACCCCGCCCGGAGACCATAGTGTCTGCGAACCAGGAGTACACCGGAA 106

DB

743 CAGAGACAGCCCTCAGACAGTGTCTGAGAGCCGTTGTCTCAGCAAG 802

OY

107 TTGGCAGACGACCGGCTCTTCTTGA 135

DB

803 GAACACAGACAGCCTGGGCAATATTTGA 831

RESULT 7

AA172136 422 bp mRNA linear EST 23-DEC-1996
LOCUS 422 bp mRNA linear EST 23-DEC-1996
DEFINITION clone IMAGE:610882 3', mRNA sequence.

ACCESSION

AA172136

VERSION

AA172136.1 GI:1751256

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo

AUTHORS

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapelli, B., Chissoe, S., Dietrich, N., Dubucq, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Maris, E., Moore, Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, D., Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Merrin, M.

TITLE

Generation and analysis of 280,000 human expressed sequence tags

JOURNAL

Genome Res. 6 (9): 807-828 (1996)

MEDLINE

97044478

COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

FEATURES

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

source

High quality sequence stop: 328.

Location/Qualifiers

1. 422

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:610882"

BASE COUNT

80 a 145 c 123 g 73 t 1 others

Query Match 14.1%; Score 34; DB 9; Length 422;
Best Local Similarity 54.5%; Pred. No. 34;
Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

OY

96 GTACACCGAATTCAGAGACGAGCCGGCTCTTGTGATCAACCGCTCAATGCGTGA 155

DB

131 GTGATGCTAGTAGTAGAGTAGAAGCGGCTCTGCGCCCTGATGCGCTCAAGCGGTC 72

OY

156 AGATTGGGCGGTCGCCCGGAGAGCTGTAACCGAGTGTGTGGTCCGGAAGCCCTT 215

DB

71 AGGTGTGTGCAAGCCACCGGCGGTGACGCGGTGGAATTAATATAGCGCTT 12

OY

216 GTG 218

DB

11 GTG 9

RESULT 8

BB628805 706 bp mRNA linear EST 26-OCT-2001
LOCUS BB628805
DEFINITION musculus CDNA clone 9630035612 5', mRNA sequence.

ACCESSION

BB628805

VERSION

BB628805.1 GI:16465143

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE

1 (bases 1 to 706)
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyata, T., Muramatsu, M. and Hayashizaki, Y.

AUTHORS

Riken House ESTs (Arakawa, T., et al. 2001)

TITLE

Unpublished (2001)

JOURNAL

Unpublished (2001)

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp

FEATURES

URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, M., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

source

Normalizaton and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10): 1617-1630 (2000)

Location/Qualifiers

wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Itawa, M., Ohara, E., Watanabe, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

FEATURES

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multiplexed sequencer. Genome Res. 10 (11): 1757-1771 (2000)

source

Konno, H., Fukushima, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara

Location/Qualifiers

```

/tissue_type="Ascites"
/cell_type="Epithelial"
/cell_line="SNU-668"
/lab_host="Top10F"

```

81 a	143 c	137 g	107 t
------	-------	-------	-------

188

180 CACTACCTACGCTACGCGAAAGGC 212

—

BE888083/
100%

ACCESSION BE888083
 CT:10344027

SOURCE human.

Homo

AUTHORS NIH-MGC <http://mgc.nhl.nih.gov/>.
ADDRESSES Institutes of Health, Mammalian Gene Collection (MGC)

Email: cgapbs-r@mail.nih.gov
Enclosure: ATCC

<http://image.lnl.gov>

Location/Qualifiers

```
/lab_host="DH10B (phage-resistant)"
```

Average insert size 2.1 kb.

63 CCGGAGAGCCATAGTGTCTCGCGAACCGGTGAGTACACCGGAATTGCCAGGACGACCG 122

b6
c07 CANNOT BE RELEASED UNDER E.O. 1.35

ACCESSION
BI765955
CT.15757533

KEYWORDS EST, human, corporate

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)

Email: cgapbs-r@nln.gov
 Department: Life Technologies, Inc.

<http://image.11n1.gov>

Location/Qualifiers

Location/Qualifiers

```
clone_lib="NIH_MGC_116"
cd host-#du108
```

pcmv-sport6; site_1: NotI; site_2: EcoRV (ncsrfca//
sources/continuous pool of 3 colonies, age 26 yo male, 49 yo

2 a 204 c 202 g 107 t

Matches	58;	Conservative	0;	Mismatches	42;	Indels	0;	Gaps	0;
0Y	63	CGGAGAGACCTAGTGTCTCCGACACCGGTGATGATACCGGAAATTCGACAGACACCG	122						
Db	243	CAGGGCATCCAGCACTGCTGTGTGACTGTGTGGGGCCACTGGGGGCTGTGGAGCTCAGG	184						
0Y	123	GTCCTTCCTTGATCAACCGGCTCAATGCGCGAGATTTG	162						
Db	183	GGCCATCACTGTGCTGCACCTCGCTCCACGCGCTGGGGGAATG	144						
RESULT 13									
LOCUS	CNS0091P	925 bp	DNA	linear	GSS 03-JUN-1999				
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.								
ACCESSION	AL053013								
VERSION	AL053013.1	GI:4934461							
KEYWORDS	GSS.								
SOURCE	Drosophila melanogaster.								
ORGANISM	Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.								
REFERENCE	1 (bases 1 to 925) Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : : BP 191 91006 Evry cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)								
AUTHORS	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .								
FEATURES	Location/Qualifiers								
SOURCE	1..925. /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACR19D16" /clone.lib="RPCI-98" /note="end : TET3"								
BASE COUNT	120 a 61 c 61 g 172 t 511 others								
ORIGIN									
Query Match	13.6% Score 32.8; DB 17; Length 925;								
Best Local Similarity	13.7% Pred. No. 91;								
Matches	29; Conservative 98; Mismatches 85; Indels 0; Gaps 0;								
0Y	1	GCAGAAACGCTAGCCATGCGCTAGTATGATGCTCGTGACGCTCCAGACCCCCCT	60						
Db	707	SSSVSSGSKSTSSSGSSSSSSSTSBABCTSTSSSSSSSYSSSTSCCTCCS	766						
0Y	61	CCCCGGAGACCATAGTGTCTCCGGAACCGGTGAGTACACCGGAATTCGACGACGACC	120						
Db	767	YSYSSTSSSTSWGSTSSGSSSSSVGSSSSSDSTSCCCTCCCTCCTYBMCCTYST	826						
0Y	121	GGGTCCTTTCTTGATCAACCGGCTCAATGCTGAGATTTGGCGGTGCCCCCGGAGAC	180						
Db	827	SCGGSSSSSGGATYKCCGCGGSSSTGMBGTSSACSSSSSSSVSSSSSKSSASS	886						
0Y	181	TGCTAGCCGAGTAGTGTGGTGGCTGGCAAGGC	212						

```

Db          : 887  SVSSSSGVSNSSS$ASKSSSSGVS$GSS 918

RESULT 14
BF306437/c
LOCUS      601893322P1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4138900 5',
DEFINITION mRNA sequence.
ACCESSION  BF306437
VERSION     BF306437.1  GI:11253532
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 1331)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL).
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: Image.lnl.gov
Plate: LCM1048 row: 1 column: 05
High quality sequence stop: 642.
Location/Qualifiers
1. 1331
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4138900"
/clone_1ib="NIH_MGC_17"
/tissue_type="rhadomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAC(G). Size selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```


Fri Jul 11 15:29:33 2003

us-10-087-631b-10.rst

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 2670)
Strausberg R.
Direct Submission
Submitted (10-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Louis Staudt
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@ngri.nih.gov
Shevchenko, Y., Wetherby, R.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantrop, S., Thomas, P.J.,
Tongson, E.E., Touchman, J.W., Tsurganov, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAL Plate: 261 Row: 1 Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis
This clone has the following problem: Incomplete processing

FEATURES
source

Location/Qualifiers
1..2670
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4309173"
/tissue_type="Primary B-Cells from Tonsils"
/clone_lib="NIH-MGC_48"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

BASE COUNT
ORIGIN

657 a 742 c 774 g 497 t

Query Match 13.6%; Score 32.8; DB 11; Length 2670;
Best Local Similarity 58.0%; Pred. No. 1.2e+02;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 63 GGGGAGGCGATAGTGTCTGGGACCGGTGAGTACACCGAAATTGCCAGACGACCGG 122
DB 776 CAGGGCATCCACACACTGTCTGTGACTCTGTAGGGCCACTGGGGTGTGAGACGTAGG 717
QY 123 GTCCCTTCTTGATCAACCCGCTCATGCTTGAGATTG 162
DB 716 GGGCATCACTGCTGCTCAGCTCCACGCGCTGGGGGAATG 677

Search completed: July 11, 2003, 02:26:19
Job time : 4340.13 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 18:15:27 ; Search time 2088.31 Seconds

(without alignments)
3358.597 Million cell updates/sec

Title: US-10-087-631B-10

Perfect score: 241
Sequence: 1 gcagaagcgtctagccatg.....ctgcctgataggtgcttgc 241

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_da:*

2: gb_hcg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sy:*

12: gb_un:*

13: gb_vl:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sy:*

28: em_un:*

29: em_vl:*

30: em_hcg_hum:*

31: em_hcg_inv:*

32: em_hcg_other:*

33: em_hcg_mus:*

34: em_hcg_pln:*

35: em_hcg_rnd:*

36: em_hcg_mam:*

37: em_hcg_vrt:*

38: em_sy:*

39: em_hcg_hum:*

40: em_hcg_mus:*

41: em_hcg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	241	100.0	244	6 A98437	A98437 Sequence 21
2	241	100.0	254	14 D31724	D31724 Hepatitis C
3	241	100.0	256	14 AF217300	AF217300 Hepatitis C
4	241	100.0	279	14 HPCBCU5	L34388 Hepatitis C
5	241	100.0	305	6 AX398187	AX398187 Sequence
6	241	100.0	321	14 HPCRND3	M84842 Hepatitis C
7	241	100.0	321	14 HPCRND5	M84841 Hepatitis C
8	241	100.0	321	14 HPCRND4	M84859 Hepatitis C
9	241	100.0	321	14 HPCRND5	M84859 Hepatitis C
10	241	100.0	321	14 HPCRND5	M84859 Hepatitis C
11	241	100.0	321	14 HPCRND5	M84859 Hepatitis C
12	241	100.0	321	14 HPCRND5	M84859 Hepatitis C
13	241	100.0	321	14 HPCRND5	M84859 Hepatitis C
14	241	100.0	321	14 HPCRND5	M84859 Hepatitis C
15	241	100.0	321	14 HPCRND5	M84859 Hepatitis C
16	241	100.0	321	14 HPCRND5	M84859 Hepatitis C
17	241	100.0	321	14 HPCRND5	M84859 Hepatitis C
18	241	100.0	321	14 HPCRND5	M84859 Hepatitis C
19	241	100.0	321	14 HPCRND5	M84859 Hepatitis C
20	241	100.0	328	6 AX419905	AX419905 Sequence
21	241	100.0	341	6 AR119855	AR119855 Sequence
22	241	100.0	341	6 AR119856	AR119856 Sequence
23	241	100.0	341	6 AR153763	AR153763 Sequence
24	241	100.0	341	14 HPC5NR2	D31602 Human hepat
25	241	100.0	341	14 HPC5NR2	L40552 Hepatitis C
26	241	100.0	342	6 AR095001	AR095001 Sequence
27	241	100.0	366	6 AX282476	AX282476 Sequence
28	241	100.0	395	14 HCU05028	U05028 Hepatitis C
29	241	100.0	483	6 E06161	E06161 cDNA encodl
30	241	100.0	483	6 E06357	E06357 cDNA encodl
31	241	100.0	587	6 E50749	E50749 Vector expr
32	241	100.0	686	6 AR142355	AR142355 Sequence
33	241	100.0	686	6 AR167008	AR167008 Sequence
34	241	100.0	686	6 AR210663	AR210663 Sequence
35	241	100.0	703	6 E50750	E50750 Vector expr
36	241	100.0	780	6 AR095006	AR095006 Sequence
37	241	100.0	866	14 HCU94722	U94722 Hepatitis C
38	241	100.0	923	6 AR153761	AR153761 Sequence
39	241	100.0	1499	6 AR005071	AR005071 Sequence
40	241	100.0	1499	6 AR064510	AR064510 Sequence
41	241	100.0	1499	6 AR091903	AR091903 Sequence
42	241	100.0	1499	6 AR145552	AR145552 Sequence
43	241	100.0	1554	6 E06166	E06166 cDNA encodl
44	241	100.0	1554	6 E06362	E06362 cDNA encodl
45	241	100.0	1580	14 DB88470	DB88470 Hepatitis C

ALIGNMENTS

RESULT 1

A98437 LOCUS A98437

DEFINITION Sequence 21 from Patent WO9912948.

ACCESSION A98437

VERSION A98437.1 GI:6781538

KEYWORDS

SOURCE

' ORGANISM

REFERENCE 1 (bases 1 to 244)

AUTHORS

TITLE

JOURNAL

Protein-coated polyribonucleic acids, method for the production thereof, and use of the same

Patent: WO 9912948-A 21 18-MAR-1999;

FEATURES
SOURCE
LANDT OLFERT (DE)
Location/Qualifiers
1..244
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT
45 a 67 c 81 g 51 t

Query Match
Best Local Similarity 100.0%; Score 241; DB 6; Length 244;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCCATGGCGTTAGTATGATGTCGTGACAGCCCTCCAGAGACCCCTT 60
DB 1 GCAGAAAGCGCTAGCCATGGCGTTAGTATGATGTCGTGACAGCCCTCCAGAGACCCCTT 60
QY 1 GCAGAAAGCGCTAGCCATGGCGTTAGTATGATGTCGTGACAGCCCTCCAGAGACCCCTT 60
DB 1 GCAGAAAGCGCTAGCCATGGCGTTAGTATGATGTCGTGACAGCCCTCCAGAGACCCCTT 60
QY 61 CCCGGAGAGCCATGATGTCGTGCGGAGAACCGGTGAGTACACCGGAATTGCCAGAGAC 120
DB 61 CCCGGAGAGCCATGATGTCGTGCGGAGAACCGGTGAGTACACCGGAATTGCCAGAGAC 120
QY 121 GGGTCTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGGGTGCCCCGGAGAC 180
DB 121 GGGTCTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGGGTGCCCCGGAGAC 180
QY 181 TCGTAGCCGAGTAGTGTGGTGGCGGAAAGGCTTGTGCTACTGCTGATAGAGGTGCTTG 240
DB 181 TCGTAGCCGAGTAGTGTGGTGGCGGAAAGGCTTGTGCTACTGCTGATAGAGGTGCTTG 240
QY 241 C 241
DB 241 C 241

RESULT 2
D31724 254 bp RNA linear VRL 09-SEP-1997
LOCUS Hepatitis C virus genomic RNA, 5'UTR, clone:BUG.
DEFINITION D31724
ACCESSION D31724
VERSION D31724.1 GI:2388524
KEYWORDS
SOURCE Hepatitis C virus cDNA to genomic RNA, clone:BUG.
ORGANISM Hepatitis C virus
VIRUSES: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE
1 (sites)
AUTHORS Onba,K., Mizokami,M., Ohno,T., Suzuki,K., Orto,E., Ina,Y.,
Lau,J.Y. and Gojobori,T.
TITLE Classification of hepatitis C virus into major types and subtypes
based on molecular evolutionary analysis
JOURNAL Virus Res. 36 (2-3), 201-214. (1995)
MEDLINE 95381614
REFERENCE 2 (bases 1 to 254)
AUTHORS Suzuki,K.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-1994) Kaoru Suzuki, Nagoya City University
Medical School, Second Department of Internal Medicine, 1-Kawasumi,
Mizuho, Mizuho-ku, Nagoya, Aichi 466, Japan (Tel:052-853-8748,
Fax:052-852-0849).

FEATURES
SOURCE
Location/Qualifiers
1..254
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
/clone="BUG"

5'UTR
BASE COUNT
46 a 72 c 82 g 54 t

Query Match
Best Local Similarity 100.0%; Score 241; DB 14; Length 254;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCCATGGCGTTAGTATGATGTCGTGACAGCCCTCCAGAGACCCCTT 60
DB 1 GCAGAAAGCGCTAGCCATGGCGTTAGTATGATGTCGTGACAGCCCTCCAGAGACCCCTT 60

DB 6 GCAGAAAGCGCTAGCCATGGCGTTAGTATGATGTCGTGACAGCCCTCCAGAGACCCCTT 65
QY 61 CCCGGAGAGCCATGATGTCGTGCGGAGAACCGGTGAGTACACCGGAATTGCCAGAGAC 120
DB 66 CCCGGAGAGCCATGATGTCGTGCGGAGAACCGGTGAGTACACCGGAATTGCCAGAGAC 125
QY 121 GGGTCTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGGGTGCCCCGGAGAC 180
DB 126 GGGTCTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGGGTGCCCCGGAGAC 185
QY 181 TCGTAGCCGAGTAGTGTGGTGGCGGAAAGGCTTGTGCTACTGCTGATAGAGGTGCTTG 240
DB 186 TCGTAGCCGAGTAGTGTGGTGGCGGAAAGGCTTGTGCTACTGCTGATAGAGGTGCTTG 245
QY 241 C 241
DB 246 C 246

RESULT 3
AF217300 256 bp RNA linear VRL 23-MAR-2000
LOCUS AF217300
DEFINITION Hepatitis C virus clone Sol12 5'UTR sequence.
ACCESSION AF217300
VERSION AF217300.1 GI:7288521
KEYWORDS
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
VIRUSES: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE
1 (bases 1 to 256)
AUTHORS Green,S.M. and Hadfield,S.J.
TITLE Direct Submission
JOURNAL Submitted (20-DEC-1999) Microbiology, Public Health Laboratory
Service, Southampton General Hospital, Southampton, Hampshire SO16
6YD, United Kingdom

FEATURES
SOURCE
Location/Qualifiers
1..256
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
/clone="Sol12"
/note="genotype: 1b"

5'UTR
BASE COUNT
46 a 73 c 85 g 52 t

Query Match
Best Local Similarity 100.0%; Score 241; DB 14; Length 256;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGCTAGCCATGGCGTTAGTATGATGTCGTGACAGCCCTCCAGAGACCCCTT 60
DB 4 GCAGAAAGCGCTAGCCATGGCGTTAGTATGATGTCGTGACAGCCCTCCAGAGACCCCTT 63
QY 61 CCCGGAGAGCCATGATGTCGTGCGGAGAACCGGTGAGTACACCGGAATTGCCAGAGAC 120
DB 64 CCCGGAGAGCCATGATGTCGTGCGGAGAACCGGTGAGTACACCGGAATTGCCAGAGAC 123
QY 121 GGGTCTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGGGTGCCCCGGAGAC 180
DB 124 GGGTCTTCTTGTGATCAACCCGCTCAATGCTGAGATTGGGGGTGCCCCGGAGAC 183
QY 181 TCGTAGCCGAGTAGTGTGGTGGCGGAAAGGCTTGTGCTACTGCTGATAGAGGTGCTTG 240
DB 184 TCGTAGCCGAGTAGTGTGGTGGCGGAAAGGCTTGTGCTACTGCTGATAGAGGTGCTTG 243
QY 241 C 241
DB 244 C 244

RESULT 4
HPCBC05

LOCUS	279 bp	DNA	linear	VRL 16-MAY-1995
DEFINITION	HepBCu5			
ACCESSION	L34388			
VERSION	L34388.1	GI:511628		
KEYWORDS	noncoding region.			
SOURCE	Hepatitis C virus DNA.			
ORGANISM	Hepatitis C virus			
REFERENCE	1 (bases 1 to 279)			
AUTHORS	Altamirano, M., Delaney, A., Wong, A., Marosteenaki, J., and Pi, D.			
TITLE	Identification of hepatitis C virus genotypes among hospitalized patients in British Columbia, Canada			
JOURNAL	J. Infect. Dis. 171 (4), 1034-1038 (1995)			
MEDLINE	95221959			
PUBMED	7706784			
FEATURES	Location/Qualifiers			
source	1..279			
BASE COUNT	49 a 79 c 92 g 59 t			
ORIGIN	/organism="Hepatitis C virus"			
	/db_xref="taxon:11103"			
Query Match	100.0%	Score 241;	DB 14;	Length 279;
Best Local Similarity	100.0%	Pred. No. 6.5e-60;		
Matches 241:	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1 GCAGAAAGCGCTCTAGCCCATGGCGTTAGTATAGTGTGCGAGCGCTCCAGAGCCCCCTT 60			
Db	9 GCAGAAAGCGCTCTAGCCCATGGCGTTAGTATAGTGTGCGAGCGCTCCAGAGCCCCCTT 68			
QY	61 CCGGGAAGAGCCATAGTGTGTGCGGAACCGGATGATACACCGGAATTGCCAGAGAGACC 120			
Db	69 CCGGGAAGAGCCATAGTGTGTGCGGAACCGGATGATACACCGGAATTGCCAGAGAGACC 128			
QY	121 GGGTCCTTTCTTGATCAACCCGCTCAATGCTTGAGATTTGGGCGTCCCGCCGAGAC 180			
Db	129 GGGTCCTTTCTTGATCAACCCGCTCAATGCTTGAGATTTGGGCGTCCCGCCGAGAC 188			
QY	181 TGTAGGCGAGTAGTGTGGTGCAGAAAGCCCTTGTGTAAGTCTGCTGATAGGGTGTG 240			
Db	189 TGTAGGCGAGTAGTGTGGTGCAGAAAGCCCTTGTGTAAGTCTGCTGATAGGGTGTG 248			
QY	241 C 241			
Db	249 C 249			
RESULT 5				
LOCUS	AX398187	305 bp	DNA	linear
DEFINITION	Sequence 64 from Patent WO220837.			
ACCESSION	AX398187			
VERSION	AX398187.1	GI:21261002		
KEYWORDS	Hepatitis C virus.			
SOURCE	Hepatitis C virus			
ORGANISM	Hepatitis C virus			
REFERENCE	1			
AUTHORS	Ronaghi, M., Ekstroem, B. and Pourmand, N.			
TITLE	Method			
JOURNAL	Patent: WO 0220837-A 64 14-MAR-2002;			
FEATURES	Pyrosequencing AB (SE) ; The Board of Trustees of The Leland Stanford Junior University (US)			
source	Location/Qualifiers			
	1..305			
	/organism="Hepatitis C virus"			
	/db_xref="taxon:11103"			
BASE COUNT	56 a 87 c 98 g 64 t			
ORIGIN				
Query Match	100.0%	Score 241;	DB 6;	Length 305;

[illegible]

	RESULT 6	
HPCRNAD3		
LOCUS	321 bp ss-RNA	linear VRL 19-APR-1994
DEFINITION	Hepatitis C virus isolate D3 5' untranslated region and polyprotein	
ACCESSION	M84842	
VERSION	M84842.1 GI:329889	
KEYWORDS	polyprotein.	
SOURCE	Hepatitis C virus (individual isolate D3) cDNA to genomic RNA.	
ORGANISM	Hepatitis C virus	
REFERENCE	Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;	
AUTHORS	Hepacivirus.	
TITLE	I (bases 1 to 321)	
JOURNAL	Bukh, D., Purcell, R.H. and Miller, R.H.	
MEDLINE	Sequence analysis of the 5' noncoding region of hepatitis C virus	
PUBMED	Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)	
FEATURES	92279243	
SOURCE	1317578	
Location/Qualifiers		
/organism="Hepatitis C virus"		
/isolate="D3"		
/db_xref="taxon:11103"		
.1..282		
283..321		
/gene="polyprotein"		
283..321		
/partial		
/gene="polyprotein"		
/codon_start=1		
/product="polyprotein"		
/protein_id="AA45679.1"		
/db_xref="GI:329890"		
/translation="MSTNPKPQRKTKR"		
BASE COUNT	69 a 91 c 97 g 64 t	
ORIGIN		
Query Match	100.0%; Score 241; DB 14; Length 321;	
Best Local Similarity	100.0%; Pred. No. 6,5e-60;	
Matches 241: Conservative 0; Mismatches .0; Indels 0; Gaps 0;		
OY	1 GCAGAAAGCGCTGTACCATGCGCTTAGTATAGTGTGTGCAGCCTCCAGAGACCCCCCT	60
DB	9 GCAAAAAGCGCTGTACCATGCGCTTAGTATAGTGTGTGCAGCCTCCAGAGACCCCCCT	68
OY	61 CCCGGAGAGCGCATATAGTGCTGTGGGAACCGGTAGTACACGGGAATTGCCAGAGACGAC	120
DB	69 CCCGGAGAGCGCATATAGTGCTGTGGGAACCGGTAGTACACGGGAATTGCCAGAGACGAC	128

QY 121 GGGTCCTTTCTTGATCAACCCGCTCATCTGAGATTGGGCGGCCCCGGAGAC 180
Db 129 GGGTCCTTTCTTGATCAACCCGCTCATCTGAGATTGGGCGGCCCCGGAGAC 188
QY 181 TGGTAGCCGAGTAGTGTGGGTGCGAAGAGCCTTGTGCTACTGCTGATAGGGTCTTG 240
Db 189 TGGTAGCCGAGTAGTGTGGGTGCGAAGAGCCTTGTGCTACTGCTGATAGGGTCTTG 248
QY 241 C 241
Db 249 C 249

RESULT 7
HPCRNAD6 321 bp ss-RNA linear VRL 19-Apr-1994
LOCUS Hepatitis C virus isolate D6 5' untranslated region and polyprotein
DEFINITION M84841.1 GI:329891
ACCESSION M84841.1 GI:329891
VERSION M84841.1 GI:329891
KEYWORDS Hepatitis C virus (individual isolate D6) cDNA to genomic RNA.
SOURCE Hepatitis C virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
REFERENCE 1 (bases 1 to 321)
AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.
TITLE Sequence analysis of the 5' noncoding region of hepatitis C virus
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
MEDLINE 92279243
PUBMED 1317578

FEATURES
source
1..321
Location/Qualifiers
/organism="Hepatitis C virus"
/isolate="D6"
/db_xref="taxon:11103"
1..282
283..321
/gene="polyprotein"
283..321
/partial
/gene="polyprotein"
/codon_start=1
/product="polyprotein"
/protein_id="AA45680.1"
/db_xref="GI:329892"
/translation="MSTNPKPQRKTR"
69 a 91 c 97 g 64 t

BASE COUNT
ORIGIN
Query Match 100.0%; Score 241; DB 14; Length 321;
Best Local Similarity 100.0%; Pred. No. 6, 5e-60;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTACGATGCGGCTTACTAGTGTGCTGACAGCTCCAGAGCCGCCCT 60
Db 9 GCAGAAAGCGTACGATGCGGCTTACTAGTGTGCTGACAGCTCCAGAGCCGCCCT 68
QY 61 CCCGGAGAGCCATAGTGTGCTGCGAAGCGGTGAGTACACCGGAATTGGCAGAGAC 120
Db 69 CCCGGAGAGCCATAGTGTGCTGCGAAGCGGTGAGTACACCGGAATTGGCAGAGAC 128
QY 121 GGGTCCTTTCTTGATCAACCCGCTCATCTGAGATTGGGCGGCCCCGGAGAC 180
Db 129 GGGTCCTTTCTTGATCAACCCGCTCATCTGAGATTGGGCGGCCCCGGAGAC 188
QY 181 TGGTAGCCGAGTAGTGTGGGTGCGAAGAGCCTTGTGCTACTGCTGATAGGGTCTTG 240
Db 189 TGGTAGCCGAGTAGTGTGGGTGCGAAGAGCCTTGTGCTACTGCTGATAGGGTCTTG 248
QY 241 C 241
Db 249 C 249

RESULT 8
HPCRNAD4 321 bp ss-RNA linear VRL 19-Apr-1994
LOCUS Hepatitis C virus isolate DR4 5' untranslated region and
DEFINITION M84859.1 GI:329897
ACCESSION M84859.1 GI:329897
VERSION M84859.1 GI:329897
KEYWORDS Hepatitis C virus (individual isolate DR4) cDNA to genomic RNA.
SOURCE Hepatitis C virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
REFERENCE 1 (bases 1 to 321)
AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.
TITLE Sequence analysis of the 5' noncoding region of hepatitis C virus.
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
MEDLINE 92279243
PUBMED 1317578

FEATURES
source
1..321
Location/Qualifiers
/organism="Hepatitis C virus"
/isolate="DR4"
/db_xref="taxon:11103"
1..282
283..321
/gene="polyprotein"
283..321
/partial
/gene="polyprotein"
/codon_start=1
/product="polyprotein"
/protein_id="AA45683.1"
/db_xref="GI:329898"
/translation="MSTNPKPQRKTR"
69 a 91 c 97 g 64 t

BASE COUNT
ORIGIN
Query Match 100.0%; Score 241; DB 14; Length 321;
Best Local Similarity 100.0%; Pred. No. 6, 5e-60;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTACGATGCGGCTTACTAGTGTGCTGACAGCTCCAGAGCCGCCCT 60
Db 9 GCAGAAAGCGTACGATGCGGCTTACTAGTGTGCTGACAGCTCCAGAGCCGCCCT 68
QY 61 CCCGGAGAGCCATAGTGTGCTGCGAAGCGGTGAGTACACCGGAATTGGCAGAGAC 120
Db 69 CCCGGAGAGCCATAGTGTGCTGCGAAGCGGTGAGTACACCGGAATTGGCAGAGAC 128
QY 121 GGGTCCTTTCTTGATCAACCCGCTCATCTGAGATTGGGCGGCCCCGGAGAC 180
Db 129 GGGTCCTTTCTTGATCAACCCGCTCATCTGAGATTGGGCGGCCCCGGAGAC 188
QY 181 TGGTAGCCGAGTAGTGTGGGTGCGAAGAGCCTTGTGCTACTGCTGATAGGGTCTTG 240
Db 189 TGGTAGCCGAGTAGTGTGGGTGCGAAGAGCCTTGTGCTACTGCTGATAGGGTCTTG 248
QY 241 C 241
Db 249 C 249

RESULT 9
HPCRNAD5 321 bp ss-RNA linear VRL 19-Apr-1994
LOCUS Hepatitis C virus isolate HK5 5' untranslated region and
DEFINITION M84849.1 GI:329901
ACCESSION M84849.1 GI:329901
VERSION M84849.1 GI:329901
KEYWORDS Hepatitis C virus (individual isolate HK5) cDNA to genomic RNA.
SOURCE

ORGANISM Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.

REFERENCE 1 (bases 1 to 321)
 AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.
 TITLE Sequence analysis of the 5' noncoding region of hepatitis C virus
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
 MEDLINE 92279243
 PUBMED 1317578

FEATURES
 source Location/Qualifiers
 1..321
 /organism="Hepatitis C virus"
 /isolate="HK5"
 /db_xref="taxon:11103"
 5'UTR 1..282
 gene /gene="polyprotein"
 283..321
 CDS /partial
 /gene="polyprotein"
 /codon_start=1
 /product="polyprotein"
 /protein_id="AA45685.1"
 /db_xref="GI:329902"
 /translation="MSTNPKPQRTKR"
 BASE COUNT 69 a 91 c 97 g 64 t
 ORIGIN

Query Match 100.0%; Score 241; DB 14; Length 321;
 Best Local Similarity 100.0%; Pred. No. 6,5e-60;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGCTTATGATGATGCTGTCGAGCCTCCAGACCCCCCT 60
 Db 9 GCAGAAAGCGTCTAGCCATGCGCTTATGATGATGCTGTCGAGCCTCCAGACCCCCCT 68
 QY 61 CCCGGAGAGCCATAGTGTCTGCGGAACCGGTAGTACACCGGAATTGCCAGAGCAGC 120
 Db 69 CCCGGAGAGCCATAGTGTCTGCGGAACCGGTAGTACACCGGAATTGCCAGAGCAGC 128
 QY 121 GGGTCCTTTCTTGATCAACCCGCTCAATGCTGAGATTGGCGTCCCGCGAGAC 180
 Db 129 GGGTCCTTTCTTGATCAACCCGCTCAATGCTGAGATTGGCGTCCCGCGAGAC 188
 QY 181 TGTAGCCGAGTACTGTGGTGGCGAAAGCGCTTGTGCTCTGCTGATAGGCTCTTG 240
 Db 189 TGTAGCCGAGTACTGTGGTGGCGAAAGCGCTTGTGCTCTGCTGATAGGCTCTTG 248
 QY 241 C 241
 Db 249 C 249

RESULT 10
 HPCRNP10 321 bp ss-RNA linear VRL 19-APR-1994
 LOCUS Hepatitis C virus isolate P10 5' untranslated region and
 DEFINITION polyprotein mRNA, 5' end.
 ACCESSION M84855
 VERSION M84855.1 GI:329903
 KEYWORDS polyprotein.
 SOURCE Hepatitis C virus (individual isolate P10) cDNA to genomic RNA.
 ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.

REFERENCE 1 (bases 1 to 321)
 AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.
 TITLE Sequence analysis of the 5' noncoding region of hepatitis C virus
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
 MEDLINE 92279243
 PUBMED 1317578

FEATURES
 source Location/Qualifiers
 1..321

ORGANISM="Hepatitis C virus"
 /isolate="P10"
 /db_xref="taxon:11103"
 5'UTR 1..282
 gene /gene="polyprotein"
 283..321
 CDS /partial
 /gene="polyprotein"
 /codon_start=1
 /product="polyprotein"
 /protein_id="AA45686.1"
 /db_xref="GI:329904"
 /translation="MSTNPKPQRTKR"
 BASE COUNT 69 a 91 c 97 g 64 t
 ORIGIN

Query Match 100.0%; Score 241; DB 14; Length 321;
 Best Local Similarity 100.0%; Pred. No. 6,5e-60;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGAAAGCGTCTAGCCATGCGCTTATGATGATGCTGTCGAGCCTCCAGACCCCCCT 60
 Db 9 GCAGAAAGCGTCTAGCCATGCGCTTATGATGATGCTGTCGAGCCTCCAGACCCCCCT 68
 QY 61 CCCGGAGAGCCATAGTGTCTGCGGAACCGGTAGTACACCGGAATTGCCAGAGCAGC 120
 Db 69 CCCGGAGAGCCATAGTGTCTGCGGAACCGGTAGTACACCGGAATTGCCAGAGCAGC 128
 QY 121 GGGTCCTTTCTTGATCAACCCGCTCAATGCTGAGATTGGCGTCCCGCGAGAC 180
 Db 129 GGGTCCTTTCTTGATCAACCCGCTCAATGCTGAGATTGGCGTCCCGCGAGAC 188
 QY 181 TGTAGCCGAGTACTGTGGTGGCGAAAGCGCTTGTGCTCTGCTGATAGGCTCTTG 240
 Db 189 TGTAGCCGAGTACTGTGGTGGCGAAAGCGCTTGTGCTCTGCTGATAGGCTCTTG 248
 QY 241 C 241
 Db 249 C 249

RESULT 11
 HPCRNAS45 321 bp ss-RNA linear VRL 19-APR-1994
 LOCUS Hepatitis C virus isolate S45 5' untranslated region and
 DEFINITION polyprotein mRNA, 5' end.
 ACCESSION M84840
 VERSION M84840.1 GI:329907
 KEYWORDS polyprotein.
 SOURCE Hepatitis C virus (individual isolate S45) cDNA to genomic RNA.
 ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.

REFERENCE 1 (bases 1 to 321)
 AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.
 TITLE Sequence analysis of the 5' noncoding region of hepatitis C virus
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
 MEDLINE 92279243
 PUBMED 1317578

FEATURES
 source Location/Qualifiers
 1..321
 /organism="Hepatitis C virus"
 /isolate="S45"
 /db_xref="taxon:11103"
 5'UTR 1..282
 gene /gene="polyprotein"
 283..321
 CDS /partial
 /gene="polyprotein"
 /codon_start=1
 /product="polyprotein"

	GCAGAAACGCTC	AACCAATGGCGT	TATTATAGAGTG	TCGTGAGCC	MCCAGAGCCCCCCC	CG
					gaps	0;
					indels	0;
					mismatches	0;
1						

121 GGGTCTTCTTGATCAACCCGCTCATTCCTGGAGATTGGCGTGCCCCGCGAGAC 180

129 GGGTCCTTCTTGGATCAACCGTCATGCTGGAGATTTGGGGTGGCCCCGGGAGAC 188

RESULT 14
LOCUS HPCRNASUS3 321 bp ss-RNA linear VRL 19-APR-1994
DEFINITION Hepatitis C virus isolate US3 5' untranslated region and
polyprotein mRNA, 5' end.
ACCESSION M84857.1 GI:329933
VERSION M84857
KEYWORDS polyprotein.
SOURCE Hepatitis C virus (individual isolate US3) cDNA to genomic RNA.
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE 1 (bases 1 to 321)
AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.
TITLE Sequence analysis of the 5' noncoding region of hepatitis C virus
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
MEDLINE 92279243
PUBMED 1317578
FEATURES
source
Location/Qualifiers
1..321
/organism="Hepatitis C virus"
/isolate="US3"
/db_xref="taxon:11103"
1..282
/gene="polyprotein"
283..321
/partial
/gene="polyprotein"
/codon_start=1
/product="polyprotein"
/protein_id="AA45699.1"
/db_xref="GI:329934"
/translation="MSTNPKPQRKTR"
BASE COUNT 69 a 91 c 97 g 64 t
ORIGIN
Query Match 100.0%; Score 241; DB 14; Length 321;
Best Local Similarity 100.0%; Pred. No. 6.5e-60;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GCAGAAAGCGCTAGCCATGCGCTTGTAGTATGATGTCGTGACGCTCCAGAGACCCCTT 60
DB 9 GCAGAAAGCGCTAGCCATGCGCTTGTAGTATGATGTCGTGACGCTCCAGAGACCCCTT 68
OY 61 CCCGGAGAGCCATAGTGTGTCGGAACCGGTGAGTACACCGAATTGCCAGAGACGAC 120
DB 69 CCCGGAGAGCCATAGTGTGTCGGAACCGGTGAGTACACCGAATTGCCAGAGACGAC 128
OY 121 GGGTCCTTTCTTGGATCAACCCGCTCAATGCTGAGATTGGCGTCCCGCGAGAC 180
DB 129 GGGTCCTTTCTTGGATCAACCCGCTCAATGCTGAGATTGGCGTCCCGCGAGAC 188
OY 181 TGCTAGCCGAGTACTGTTGGTCCGGAAGGCTTGTGTACTGCTGATAGGTTGCTTG 240
DB 189 TGCTAGCCGAGTACTGTTGGTCCGGAAGGCTTGTGTACTGCTGATAGGTTGCTTG 248
OY 241 C 241
DB 249 C 249

RESULT 15
HPCRNASUS6

LOCUS HPCRNASUS6 321 bp ss-RNA linear VRL 19-APR-1994
DEFINITION Hepatitis C virus isolate US6 5' untranslated region and
polyprotein, 5' end.
ACCESSION M84830.1 GI:329935
VERSION M84830
KEYWORDS polyprotein.
SOURCE Hepatitis C virus (individual isolate US6) cDNA to genomic RNA.
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnavirus.
REFERENCE 1 (bases 1 to 321)
AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.
TITLE Sequence analysis of the 5' noncoding region of hepatitis C virus
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
MEDLINE 92279243
PUBMED 1317578
FEATURES
source
Location/Qualifiers
1..321
/organism="Hepatitis C virus"
/isolate="US6"
/db_xref="taxon:11103"
1..282
/gene="polyprotein"
283..321
/partial
/gene="polyprotein"
/codon_start=1
/product="polyprotein"
/protein_id="AA45700.1"
/db_xref="GI:329936"
/translation="MSTNPKPQRKTR"
BASE COUNT 69 a 91 c 97 g 64 t
ORIGIN
Query Match 100.0%; Score 241; DB 14; Length 321;
Best Local Similarity 100.0%; Pred. No. 6.5e-60;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GCAGAAAGCGCTAGCCATGCGCTTGTAGTATGATGTCGTGACGCTCCAGAGACCCCTT 60
DB 9 GCAGAAAGCGCTAGCCATGCGCTTGTAGTATGATGTCGTGACGCTCCAGAGACCCCTT 68
OY 61 CCCGGAGAGCCATAGTGTGTCGGAACCGGTGAGTACACCGAATTGCCAGAGACGAC 120
DB 69 CCCGGAGAGCCATAGTGTGTCGGAACCGGTGAGTACACCGAATTGCCAGAGACGAC 128
OY 121 GGGTCCTTTCTTGGATCAACCCGCTCAATGCTGAGATTGGCGTCCCGCGAGAC 180
DB 129 GGGTCCTTTCTTGGATCAACCCGCTCAATGCTGAGATTGGCGTCCCGCGAGAC 188
OY 181 TGCTAGCCGAGTACTGTTGGTCCGGAAGGCTTGTGTACTGCTGATAGGTTGCTTG 240
DB 189 TGCTAGCCGAGTACTGTTGGTCCGGAAGGCTTGTGTACTGCTGATAGGTTGCTTG 248
OY 241 C 241
DB 249 C 249

Search completed: July 10, 2003, 19:20:43
Job time : 2092.31 secs

